

9-638-693

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 11:17:59 ; Search time 33 Seconds

(without alignments)
639.716 Million cell updates/sec

Title: US-09-638-693-36

Perfect score: 133

Sequence: 1 QNEICLTHPIKTYIMACMSA.....VIEPIVTTNQKLEAFNHKH 133

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_15Jun03.*

1: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	133	100.0	133	15	AA63288 Polypeptide encode
2	98	73.7	133	15	AA63289 Polypeptide encode
3	98	73.7	133	15	AA63290 Polypeptide encode
4	98	73.7	209	15	AA63390 HCV polypeptide se
5	88	66.2	128	14	AA637932 HCV NS4 region fro
6	88	66.2	128	14	AA637937 HCV NS4 region con
7	55	41.4	117	14	AA637934 HCV NS4 region fro
8	52	39.1	127	14	AA637933 HCV NS4 region fro
9	52	39.1	133	15	AA63286 Polypeptide encode

10	52	39.1	133	15	AA63287 Polypeptide encode
11	48	36.1	117	14	AA637935 HCV NS4 region fro
12	48	36.1	127	14	AA637936 HCV NS4 region fro
13	41	30.8	829	23	AAE18690 Multiple epitope f
14	41	30.8	1099	23	AAU76378 HCV multiple epito
15	41	30.8	1099	24	ABG72262 HCV multiple epito
16	30	22.6	3023	17	AA94462 Hepatitis C virus
17	20	15.0	20	15	AA63316 Hepatitis C virus
18	20	15.0	22	14	AA41164 Polypeptide fragme
19	19	14.3	19	14	AA637941 HCV (type 3) pepti
20	19	14.3	95	15	AA63285 HCV NS-4 type 3 re
21	19	14.3	194	15	AA63285 Polypeptide encode
22	19	14.3	195	13	AA63285 HCV antigen. Synt
23	19	14.3	195	14	AA63285 HCV polypeptide 10
24	19	14.3	222	13	AA63285 Hepatitis C virus
25	19	14.3	293	17	AA63285 HK10. Hepatitis C
26	19	14.3	481	15	AA63436 HCV NS4 derived an
27	19	14.3	489	15	AA63377 HCV polypeptide se
28	19	14.3	615	19	AA637807 Hepatitis C virus
29	19	14.3	631	16	AA63285 Nonstructural doma
30	19	14.3	632	20	AA63285 NS3 serine proteas
31	19	14.3	646	20	AA63285 Native HCV NS3 pro
32	19	14.3	646	20	AA63285 HCV NS4A-NS3 compl
33	19	14.3	646	20	AA63285 HCV NS4A-NS3 compl
34	19	14.3	665	20	AA63285 HCV NS4A-NS3 compl
35	19	14.3	665	20	AA63285 HCV NS4A-NS3 compl
36	19	14.3	665	20	AA63285 HCV NS4A-NS3 compl
37	19	14.3	665	20	AA63285 HCV NS4A-NS3 compl
38	19	14.3	665	20	AA63285 HCV NS4A-NS3 compl
39	19	14.3	665	20	AA63285 HCV NS4A-NS3 compl
40	19	14.3	665	20	AA63285 HCV NS4A-NS3 compl
41	19	14.3	665	20	AA63285 HCV NS4A-NS3 compl
42	19	14.3	667	20	AA63285 HCV NS4A-NS3 compl
43	19	14.3	667	20	AA63285 HCV NS4A-NS3 compl
44	19	14.3	671	20	AA63285 HCV NS4A-NS3 compl
45	19	14.3	671	20	AA63285 HCV NS4A-NS3 compl

ALIGNMENTS

RESULT 1
AA63288
ID AA63288 standard; Protein; 133 AA.
XX
AC AA63288;
XX
DT 25-MAR-2003 (updated)
DT 01-AUG-1995 (first entry)
XX
DE Polypeptide encoded by hepatitis C virus NS3/NS4 sequence.
XX
XX Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
KW classification; immunisation; prophylaxis; serotyping.
XX
OS Hepatitis C virus type 3.
XX
XX WO9425601-A2.
XX
XX 10-NOV-1994.
XX
XX 27-APR-1994; 94WO-EP01323.
XX
XX 27-APR-1993; 93EP-0401099.
XX 05-AUG-1993; 93EP-0402019.
XX (INNO-) INNOGENETICS NV SA.
XX
XX Maertens G, Stuyver L;
XX WPI; 1994-358277/44.
XX N-PSDB; AAQ78040.
XX

PT New polynucleotide sequences from hepatitis C virus - and related
PT vectors, polypeptide(s) and antibodies, useful for immunisation,
PT treatment, diagnosis and typing of HCV isolates
PS Claim 11; Page 125; 404pp; English.
XX
CC Compositions comprising at least 5, and pref. 8 or more contiguous
CC nucleotides selected from an HCV type 3 genomic sequence, more
CC particularly (i) the region spanning positions 417-957 of the
CC Core/E1 region of HCV subtype 3a; (ii) the region spanning positions
CC 4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
CC positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
CC region spanning positions 8023-8235 of the NS5 region of the BR36
CC subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
CC sequence, may be used as primers to amplify nucleic acid from an
CC isolate belonging to a specific genotype, or as a probe for specific
CC detection/classification of nucleic acid. Polypeptides encoded by
CC the nucleotides in such compositions may be used for immunisation
CC against HCV, for the detection of antibodies directed against HCV
CC and for serotyping. This sequence corresponds to the NS3/NS4
CC region of HCV subtype 3a and is taken from a clone designated
CC BR36-20-164.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 133 AA;

Query Match 100.0%; Score 133; DB 15; Length 133;
Best Local Similarity 100.0%; Pred. No. 1.6e-118;
Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ONEICLTHPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
|||||
DB 1 ONEICLTHPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
|||||

QY 61 PAIVPDKEVLYQOYDEMECECSQAAPYTEQAQVIAHQPKGKVLGLQRTAQQAIVIEPIVT 120
|||||
DB 61 PAIVPDKEVLYQOYDEMECECSQAAPYTEQAQVIAHQPKGKVLGLQRTAQQAIVIEPIVT 120
|||||

QY 121 TNMOKLEAFWHKH 133
|||||
DB 121 TNMOKLEAFWHKH 133
|||||

RESULT 2
AAR63289
ID AAR63289 standard; Protein; 133 AA.
XX
AC AAR63289;
XX
DT 25-MAR-2003 (updated)
DT 01-AUG-1995 (first entry)
XX
XX Polypeptide encoded by hepatitis C virus NS3/NS4 sequence.
XX
XX Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
XX classification; immunisation; prophylaxis; serotyping.
XX
XX Hepatitis C virus type 3.
XX
XX WO9425601-A2.
XX
XX 10-NOV-1994.
XX
XX 27-APR-1994; 94WO-EP01323.
XX
XX 27-APR-1993; 93EP-0401099.
XX
XX 05-AUG-1993; 93EP-0402019.
XX
XX (INNO-) INNOGENETICS NV SA.
XX
XX Maertens G, Stuyver L;
XX WPI; 1994-358277/44.
XX

DR N-PSDB; AAQ78041.
XX
XX New polynucleotide sequences from hepatitis C virus - and related
XX vectors, polypeptide(s) and antibodies, useful for immunisation,
XX treatment, diagnosis and typing of HCV isolates
PS Claim 11; Page 127; 404pp; English.
XX
CC Compositions comprising at least 5, and pref. 8 or more contiguous
CC nucleotides selected from an HCV type 3 genomic sequence, more
CC particularly (i) the region spanning positions 417-957 of the
CC Core/E1 region of HCV subtype 3a; (ii) the region spanning positions
CC 4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
CC positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
CC region spanning positions 8023-8235 of the NS5 region of the BR36
CC subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
CC sequence, may be used as primers to amplify nucleic acid from an
CC isolate belonging to a specific genotype, or as a probe for specific
CC detection/classification of nucleic acid. Polypeptides encoded by
CC the nucleotides in such compositions may be used for immunisation
CC against HCV, for the detection of antibodies directed against HCV
CC and for serotyping. This sequence corresponds to the NS3/NS4
CC region of HCV subtype 3a and is taken from a clone designated
CC BR36-20-166.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 133 AA;

Query Match 73.7%; Score 98; DB 15; Length 133;
Best Local Similarity 100.0%; Pred. No. 2.9e-85;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ONEICLTHPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
|||||
DB 1 ONEICLTHPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
|||||

QY 61 PAIVPDKEVLYQOYDEMECECSQAAPYTEQAQVIAHQPK 98
|||||
DB 61 PAIVPDKEVLYQOYDEMECECSQAAPYTEQAQVIAHQPK 98
|||||

RESULT 3
AAR63290
ID AAR63290 standard; Protein; 133 AA.
XX
AC AAR63290;
XX
DT 25-MAR-2003 (updated)
DT 01-AUG-1995 (first entry)
XX
XX Polypeptide encoded by hepatitis C virus NS3/NS4 sequence.
XX
XX Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
XX classification; immunisation; prophylaxis; serotyping.
XX
XX Hepatitis C virus type 3.
XX
XX WO9425601-A2.
XX
XX 10-NOV-1994.
XX
XX 27-APR-1994; 94WO-EP01323.
XX
XX 27-APR-1993; 93EP-0401099.
XX
XX 05-AUG-1993; 93EP-0402019.
XX
XX (INNO-) INNOGENETICS NV SA.
XX
XX Maertens G, Stuyver L;
XX WPI; 1994-358277/44.
XX
XX N-PSDB; AAQ78042.
XX

PT New polynucleotide sequences from hepatitis C virus - and related
PT vectors, polypeptide(s) and antibodies, useful for immunisation,
PT treatment, diagnosis and typing of HCV isolates
XX
PS Claim 11; Page 128-129; 404pp; English.
XX
CC Compositions comprising at least 5, and pref. 8 or more contiguous
CC nucleotides selected from an HCV type 3 genomic sequence, more
CC particularly (i) the region spanning positions 417-957 of the
CC Core/EL region of HCV subtype 3a; (ii) the region spanning positions
CC 4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
CC positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
CC region spanning positions 8023-8235 of the NS5 region of the BR36
CC subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
CC sequence, may be used as primers to amplify nucleic acid from an
CC isolate belonging to a specific genotype, or as a probe for specific
CC detection/classification of nucleic acid. Polypeptides encoded by
CC the nucleotides in such compositions may be used for immunisation
CC against HCV, for the detection of antibodies directed against HCV
CC and for serotyping. This sequence corresponds to the NS3/NS4
CC region of HCV subtype 3a and is taken from a clone designated
CC BR36-20-165.
CC (Updated on 25-MAR-2003 to correct PN field.)
CC
SQ Sequence 133 AA;

Query Match 73.7%; Score 98; DB 15; Length 133;
Best Local Similarity 100.0%; Pred. No. 2.9e-85;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QNEICLTHPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSVGCVVIVGHIELGSK 60
DB 1 QNEICLTHPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSVGCVVIVGHIELGSK 60

QY 61 PAIVDPKEVLYQYDEMEECSQAAPYIEQAQVIAHQFK 98
DB 61 PAIVDPKEVLYQYDEMEECSQAAPYIEQAQVIAHQFK 98

RESULT 4
AAR63390
ID AAR63390 standard; Protein; 209 AA.
XX
AC AAR63390;
XX
DT 25-MAR-2003 (updated)
DT 18-AUG-1995 (first entry)
XX
DE HCV polypeptide sequence.
XX
KW Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
KW classification; immunisation; prophylaxis; serotyping.
XX
OS Hepatitis C virus.
XX
PN W09425601-A2.
XX
PD 10-NOV-1994.
XX
PF 27-APR-1994; 94WO-EP01323.
XX
PR 27-APR-1993; 93EP-0401099.
PR 05-AUG-1993; 93EP-0402019.
XX
PA (INNO-) INNOGENETICS NV SA.
XX
PI Maertens G, Stuyver L;
XX
DR WPI; 1994-358277/44.
DR N-PSDB; AAQ78125.
XX
PT New polynucleotide sequences from hepatitis C virus - and related
PT vectors, polypeptide(s) and antibodies, useful for immunisation,
XX

PT treatment, diagnosis and typing of HCV isolates
XX
PS Disclosure; Page 274-275; 404pp; English.
XX
CC Compositions comprising at least 5, and pref. 8 or more contiguous
CC nucleotides selected from an HCV type 3 genomic sequence, more
CC particularly (i) the region spanning positions 417-957 of the
CC Core/EL region of HCV subtype 3a; (ii) the region spanning positions
CC 4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
CC positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
CC region spanning positions 8023-8235 of the NS5 region of the BR36
CC subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
CC sequence, may be used as primers to amplify nucleic acid from an
CC isolate belonging to a specific genotype, or as a probe for specific
CC detection/classification of nucleic acid. Polypeptides encoded by
CC the nucleotides in such compositions may be used for immunisation
CC against HCV, for the detection of antibodies directed against HCV
CC and for serotyping.
CC (Updated on 25-MAR-2003 to correct PN field.)
CC
SQ Sequence 209 AA;

Query Match 73.7%; Score 98; DB 15; Length 209;
Best Local Similarity 100.0%; Pred. No. 4.3e-85;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QNEICLTHPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSVGCVVIVGHIELGSK 60
DB 77 QNEICLTHPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSVGCVVIVGHIELGSK 136

QY 61 PAIVDPKEVLYQYDEMEECSQAAPYIEQAQVIAHQFK 98
DB 137 PAIVDPKEVLYQYDEMEECSQAAPYIEQAQVIAHQFK 174

RESULT 5
AAR37932
ID AAR37932 standard; Protein; 128 AA.
XX
AC AAR37932;
XX
DT 25-MAR-2003 (updated)
DT 23-SEP-1993 (first entry)
XX
DE HCV NS4 region from donor T0040.
XX
KW Non-coding region; hepatitis C virus; blood donor; type 2; type 1;
KW HCV; NS-5; phylogeny; differentiation; NS-3; core region; type 3.
XX
OS Hepatitis C virus.
XX
PN W09310239-A2.
XX
PD 27-MAY-1993.
XX
PF 20-NOV-1992; 92WO-GB02143.
XX
PR 21-NOV-1991; 91GB-0024696.
PR 24-JUN-1992; 92GB-0013362.
XX
PA (COMM-) COMMON SERVICES AGENCY.
XX
PI Chan S, Simmonds P, Yap PL;
XX
DR WPI; 1993-182554/22.
DR N-PSDB; AAQ43106.
XX
PT DNA encoding antigenic peptide(s) of new types of hepatitis C
PT virus - for diagnosing and treating HCV infection, screening
PT blood samples and identifying different HCV types
XX
PS Disclosure; Fig 9b; 120pp; English.
XX

CC The sequences given in AAR37932-37 show amino acids 1638-1765 of the NS4
 CC region of hepatitis C virus-3 (HCV-3) samples from 5 blood donors
 CC and a consensus sequence. Analysis of this and other regions of the
 CC HCV genome revealed the existence of three distinct groups of HCV.
 CC Analysis of the region encompassing -255 to -62 of the 5' non coding
 CC region (see AAQ43058-75) showed a difference of 9-14% in the nucleotide
 CC sequences between the three groups. Two of the groups identified were
 CC similar to those of HCV variants termed type 1 and 2, whilst the third
 CC region (see AAR37927-30) showed a high degree of sequence diversity with
 CC type 3 being phylogenetically different to type 1 and 2. The same
 CC degree of differentiation was noted in the NS-5 (see AAR37923-26) and core
 CC region between type 3 and type 1 sequences.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 128 AA;

Query Match 66.2%; Score 88; DB 14; Length 128;
 Best Local Similarity 100.0%; Pred. No. 9.1e-76;
 Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 TKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSVGCVIVGHIELGGKPAIYVPDKVL 70
 |||||
 DB 5 TKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSVGCVIVGHIELGGKPAIYVPDKVL 64

OY 71 YQYDEMEECSSQAAPYIEQAQVIAHQFK 98
 |||||
 DB 65 YQYDEMEECSSQAAPYIEQAQVIAHQFK 92

RESULT 6

AAR37937

ID AAR37937 standard; protein; 128 AA.

XX

AC AAR37937;

XX

DT 25-MAR-2003 (updated)

DT 23-SEP-1993 (first entry)

XX

DE HCV NS4 region consensus sequence.

XX

KW Non-coding region; hepatitis C virus; blood donor; type 2; type 1;

KW HCV; NS-5; phylogeny; differentiation; NS-3; core region; type 3.

XX

OS Synthetic.

XX

XX WO9310239-A2.

XX

PD 27-MAY-1993.

XX

XX 20-NOV-1992; 92WO-GB02143.

XX

XX 21-NOV-1991; 91GB-0024696.

PR

PR 24-JUN-1992; 92GB-0013362.

XX

XX (COMM-) COMMON SERVICES AGENCY.

XX

XX Chan S, Simmonds P, Yap PL;

XX

XX WPI; 1993-182554/22.

XX

XX N-PSDB; AAQ43111.

XX

PT DNA encoding antigenic peptide(s) of new types of hepatitis C

PT virus - for diagnosing and treating HCV infection, screening

PT

PT blood samples and identifying different HCV types

PS

PS Disclosure; Fig 9b; 120pp; English.

XX

CC The sequences given in AAR37932-37 show amino acids 1638-1765 of the NS4

CC region of hepatitis C virus-3 (HCV-3) samples from 5 blood donors

CC and a consensus sequence. Analysis of this and other regions of the

CC HCV genome revealed the existence of three distinct groups of HCV.

CC Analysis of the region encompassing -255 to -62 of the 5' non coding

CC

CC

CC

CC

CC region (see AAQ43058-75) showed a difference of 9-14% in the nucleotide
 CC sequences between the three groups. Two of the groups identified were
 CC similar to those of HCV variants termed type 1 and 2, whilst the third
 CC appeared to represent a novel type of virus. Comparison of the NS3
 CC region (see AAR37927-30) showed a high degree of sequence diversity with
 CC type 3 being phylogenetically different to type 1 and 2. The same
 CC degree of differentiation was noted in the NS-5 (see AAR37923-26) and core
 CC region between type 3 and type 1 sequences.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX

SQ Sequence 128 AA;

Query Match 66.2%; Score 88; DB 14; Length 128;

Best Local Similarity 100.0%; Pred. No. 9.1e-76;

Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 TKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSVGCVIVGHIELGGKPAIYVPDKVL 70
 |||||
 DB 5 TKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSVGCVIVGHIELGGKPAIYVPDKVL 64

OY 71 YQYDEMEECSSQAAPYIEQAQVIAHQFK 98

|||||

DB 65 YQYDEMEECSSQAAPYIEQAQVIAHQFK 92

RESULT 7

AAR37934

ID AAR37934 standard; protein; 117 AA.

XX

AC AAR37934;

XX

DT 25-MAR-2003 (updated)

DT 23-SEP-1993 (first entry)

XX

DE HCV NS4 region from donor T0036.

XX

KW Non-coding region; hepatitis C virus; blood donor; type 2; type 1;

KW HCV; NS-5; phylogeny; differentiation; NS-3; core region; type 3.

XX

OS Hepatitis C virus.

XX

PN WO9310239-A2.

XX

XX 27-MAY-1993.

XX

XX 20-NOV-1992; 92WO-GB02143.

XX

XX 21-NOV-1991; 91GB-0024696.

PR

PR 24-JUN-1992; 92GB-0013362.

XX

XX (COMM-) COMMON SERVICES AGENCY.

XX

XX Chan S, Simmonds P, Yap PL;

XX

XX WPI; 1993-182554/22.

XX

XX N-PSDB; AAQ43108.

XX

PT DNA encoding antigenic peptide(s) of new types of hepatitis C

PT virus - for diagnosing and treating HCV infection, screening

PT

PT blood samples and identifying different HCV types

XX

PS Disclosure; Fig 9b; 120pp; English.

XX

CC The sequences given in AAR37932-37 show amino acids 1638-1765 of the NS4

CC region of hepatitis C virus-3 (HCV-3) samples from 5 blood donors

CC and a consensus sequence. Analysis of this and other regions of the

CC HCV genome revealed the existence of three distinct groups of HCV.

CC Analysis of the region encompassing -255 to -62 of the 5' non coding

CC region (see AAQ43058-75) showed a difference of 9-14% in the nucleotide

CC sequences between the three groups. Two of the groups identified were

CC similar to those of HCV variants termed type 1 and 2, whilst the third

CC appeared to represent a novel type of virus. Comparison of the NS3

CC region (see AAR37927-30) showed a high degree of sequence diversity with

CC

CC

CC

CC

CC

CC type 3 being phylogenetically different to type 1 and 2. The same
 CC degree differentiation was noted in the NS-5 (see AAR37923-26) and core
 CC region between type 3 and type 1 sequences.
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 117 AA;

Query Match 41.4%; Score 55; DB 14; Length 117;
 Best Local Similarity 100.0%; Pred. No. 2e-44;
 Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 HPITKYMCMASADLEVTSTWLLGGVLAALAAAYCLSGVCVIVGHIELGKPA 62
 |||||
 DB 1 HPITKYMCMASADLEVTSTWLLGGVLAALAAAYCLSGVCVIVGHIELGKPA 55
 |||||

RESULT 8

AAR37933
 ID AAR37933 standard; Protein; 127 AA.

XX AC AAR37933;

XX 25-MAR-2003 (updated)

DT 23-SEP-1993 (first entry)

XX HCV NS4 region from donor T0038.

XX Non-coding region; hepatitis C virus; blood donor; type 2; type 1;
 KW HCV; NS-5; phylogeny; differentiation; NS-3; core region; type 3.

XX Hepatitis C virus.

XX WO9310239-A2.

XX 27-MAY-1993.

XX 20-NOV-1992; 92WO-GB02143.

XX 21-NOV-1991; 91GB-0024696.

PR 24-JUN-1992; 92GB-0013362.

XX (COMM-) COMMON SERVICES AGENCY.

XX Chan S, Simmonds P, Yap PL;

XX WPI; 1993-182554/22.

DR N-PSDB; AAQ43107.

XX DNA encoding antigenic peptide(s) of new types of hepatitis C
 PT virus - for diagnosing and treating HCV infection, screening
 PT blood samples and identifying different HCV types

XX Disclosure; Fig 9b; 120pp; English.

XX The sequences given in AAR37923-37 show amino acids 1638-1765 of the NS4
 CC region of hepatitis C virus-3 (HCV-3) samples from 5 blood donors
 CC and a consensus sequence. Analysis of this and other regions of the
 CC HCV genome revealed the existence of three distinct groups of HCV.
 CC Analysis of the region encompassing -255 to -62 of the 5' non coding
 CC region (see AAQ43058-75) showed a difference of 9-14% in the nucleotide
 CC sequences between the three groups. Two of the groups identified were
 CC similar to those of HCV variants termed type 1 and 2, whilst the third
 CC appeared to represent a novel type of virus. Comparison of the NS3
 CC region (see AAR37927-30) showed a high degree of sequence diversity with
 CC type 3 being phylogenetically different to type 1 and 2. The same
 CC degree differentiation was noted in the NS-5 (see AAR37923-26) and core
 CC region between type 3 and type 1 sequences.
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 127 AA;

Query Match 39.1%; Score 52; DB 14; Length 127;
 Best Local Similarity 100.0%; Pred. No. 1.5e-41;

Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 11 TKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSGVCVIVGHIELGKPA 62
 |||||
 DB 4 TRYIMACMSADLEVTSTWLLGGVLAALAAAYCLSGVCVIVGHIELGKPA 55
 |||||

RESULT 9

AAR63286
 ID AAR63286 standard; Protein; 133 AA.

XX AC AAR63286;

XX 25-MAR-2003 (updated)

DT 01-AUG-1995 (first entry)

XX Polypeptide encoded by hepatitis C virus NS3/NS4 sequence.

XX Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
 KW classification; immunisation; prophylaxis; serotyping.

XX Hepatitis C virus type 3.

XX WO9425601-A2.

XX 10-NOV-1994.

XX 27-APR-1994; 94WO-EP01323.

XX 27-APR-1993; 93EP-0401099.

PR 05-AUG-1993; 93EP-0402019.

XX (INNO-) INNOGENETICS NV SA.

XX Maertens G, Stuyver L;

XX WPI; 1994-358277/44.

DR N-PSDB; AAQ78038.

XX New polynucleotide sequences from hepatitis C virus - and related
 PT vectors, polypeptide(s) and antibodies, useful for immunisation,
 PT treatment, diagnosis and typing of HCV isolates

XX Claim 11; Page 121-122; 404pp; English.

XX Compositions comprising at least 5, and pref. 8 or more contiguous
 CC nucleotides selected from an HCV type 3 genomic sequence, more
 CC particularly (i) the region spanning positions 417-957 of the
 CC Core/E1 region of HCV subtype 3a; (ii) the region spanning positions
 CC 4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
 CC positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
 CC region spanning positions 8023-8235 of the NS5 region of the BR36
 CC subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
 CC sequence, may be used as primers to amplify nucleic acid from an
 CC isolate belonging to a specific genotype, or as a probe for specific
 CC detection/classification of nucleic acid. Polypeptides encoded by
 CC the nucleotides in such compositions may be used for immunisation
 CC against HCV, for the detection of antibodies directed against HCV
 CC and for serotyping. This sequence corresponds to the NS3/NS4
 CC region of HCV subtype 3a and is taken from a clone designated
 CC HD10-1-25.
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 133 AA;

Query Match 39.1%; Score 52; DB 15; Length 133;
 Best Local Similarity 100.0%; Pred. No. 1.6e-41;

Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 TKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSGVCVIVGHIELGKPA 62
 |||||

DB 11 TRYIMACMSADLEVTSTWLLGGVLAALAAAYCLSGVCVIVGHIELGKPA 62
 |||||

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 11:22:55 ; Search time 18 seconds
(without alignments)
312.630 Million cell updates/sec

Title: US-09-638-693-36
Perfect score: 133
Sequence: 1 QNEICLTHPIKYMCMAS.....VIEPIVTTNQKLEAFWKH 133

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	36.1	128	1 US-08-244-116B-17	Sequence 17, Appl
2	20	15.0	22	2 US-08-146-028-107	Sequence 107, App
3	20	15.0	22	3 US-08-723-425A-107	Sequence 107, App
4	20	15.0	22	3 US-09-112-206-107	Sequence 107, App
5	19	14.3	19	1 US-08-244-116B-2	Sequence 2, Appli
6	19	14.3	247	1 US-08-324-977-44	Sequence 44, Appl
7	19	14.3	247	2 US-08-384-616-44	Sequence 44, Appl
8	19	14.3	247	2 US-08-904-686A-44	Sequence 44, Appl
9	19	14.3	247	3 US-09-315-850-44	Sequence 44, Appl
10	19	14.3	631	1 US-08-700-356-1	Sequence 1, Appli
11	19	14.3	631	2 US-08-936-865-1	Sequence 1, Appli
12	19	14.3	632	3 US-09-198-723A-23	Sequence 23, Appl
13	19	14.3	646	3 US-09-198-723A-60	Sequence 60, Appl
14	19	14.3	646	3 US-09-198-723A-63	Sequence 63, Appl
15	19	14.3	646	3 US-09-198-723A-66	Sequence 66, Appl
16	19	14.3	646	3 US-09-198-723A-69	Sequence 69, Appl
17	19	14.3	646	3 US-09-198-723A-72	Sequence 72, Appl
18	19	14.3	665	4 US-09-543-376B-1	Sequence 1, Appli
19	19	14.3	665	4 US-09-543-376B-2	Sequence 2, Appli
20	19	14.3	665	4 US-09-543-376B-3	Sequence 3, Appli
21	19	14.3	666	3 US-09-198-723A-11	Sequence 11, Appl
22	19	14.3	666	3 US-09-198-723A-12	Sequence 12, Appl
23	19	14.3	666	3 US-09-198-723A-13	Sequence 13, Appl
24	19	14.3	666	3 US-09-198-723A-14	Sequence 14, Appl
25	19	14.3	666	3 US-09-198-723A-15	Sequence 15, Appl
26	19	14.3	666	3 US-09-198-723A-16	Sequence 16, Appl
27	19	14.3	666	3 US-09-198-723A-17	Sequence 17, Appl

28	19	14.3	666	3 US-09-198-723A-18	Sequence 18, Appl
29	19	14.3	672	3 US-09-198-723A-19	Sequence 19, Appl
30	19	14.3	672	3 US-09-198-723A-20	Sequence 20, Appl
31	19	14.3	1892	3 US-09-263-933-4	Sequence 4, Appli
32	19	14.3	1892	3 US-09-263-933-11	Sequence 11, Appl
33	19	14.3	1692	3 US-09-263-933-18	Sequence 18, Appl
34	19	14.3	2013	1 US-08-324-977-12	Sequence 12, Appl
35	19	14.3	2013	2 US-08-384-616-12	Sequence 12, Appl
36	19	14.3	2013	2 US-08-904-686A-12	Sequence 12, Appl
37	19	14.3	2013	3 US-09-315-850-12	Sequence 12, Appl
38	19	14.3	2201	4 US-08-952-981A-2	Sequence 2, Appli
39	19	14.3	2307	3 US-09-263-933-9	Sequence 9, Appli
40	19	14.3	2307	3 US-09-263-933-16	Sequence 16, Appl
41	19	14.3	2307	3 US-08-324-977-32	Sequence 32, Appl
42	19	14.3	2620	1 US-08-384-616-32	Sequence 32, Appl
43	19	14.3	2620	2 US-08-904-686A-32	Sequence 32, Appl
44	19	14.3	2620	3 US-09-315-850-32	Sequence 32, Appl
45	19	14.3	2620	3 US-09-315-850-32	Sequence 32, Appl

ALIGNMENTS

RESULT 1
US-08-244-116B-17
; Sequence 17, Application US/08244116B
; Patent No. 5763159
; GENERAL INFORMATION:
; APPLICANT: Simmonds, Peter
; APPLICANT: Chan, Shiu-Wan
; APPLICANT: Yap, Peng L.
; TITLE OF INVENTION: Hepatitis-C Virus Testing
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell, Seltzer, Park & Gibson, P.A.
; STREET: 1211 East Morehead Street
; CITY: Charlotte
; STATE: No. 5763159th Carolina
; COUNTRY: United States
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0. Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/244,116B
; FILING DATE: 15-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB92/02143
; FILING DATE: 20-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 1749-125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 704-377-1561
; TELEFAX: 704-334-2014
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 128 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: yes
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Hepatitis-C virus
US-08-244-116B-17

Query Match 36.1%; Score 48; DB 1; Length 128;

Best Local Similarity 100.0%; Pred. No. 2.7e-39;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 MACMSADLEVTSTWLLGGVLAALAAAYCLSVGCVIVGHLELGGKPA 62
Db 9 MACMSADLEVTSTWLLGGVLAALAAAYCLSVGCVIVGHLELGGKPA 56

RESULT 2
US-08-146-028-107
; Sequence 107, Application US/08146028
; Patent No. 5891840
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: PROCESS FOR THE DETERMINATION OF PEPTIDES
; TITLE OF INVENTION: CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT EPITOPES AND THEIR
; TITLE OF INVENTION: IN A PROCESS FOR DETERMINATION OF ANTIBODIES OR BIOTINYLATED
; TITLE OF INVENTION: PEPTIDES CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT EPITOPES,
; TITLE OF INVENTION: PROCESS FOR PREPARING THEM AND COMPOSITIONS CONTAINING THEM
; NUMBER OF SEQUENCES: 453
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/146.028
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-146-028-107

Query Match 15.0%; Score 20; DB 2; Length 22;
Best Local Similarity 100.0%; Pred. No. 9.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 LGGKPAIVDPKVELYQQYDE 76
Db 2 LGGKPAIVDPKVELYQQYDE 21

RESULT 3
US-08-723-425A-107
; Sequence 107, Application US/08723425A
; Patent No. 6165730
; GENERAL INFORMATION:
; APPLICANT: DELEYS, ROBERT
; TITLE OF INVENTION: PROCESS FOR THE DETERMINATION OF
; TITLE OF INVENTION: PEPTIDES CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT
; TITLE OF INVENTION: EPITOPES AND THEIR USE IN A PROCESS FOR DETERMINATION OF
; TITLE OF INVENTION: ANTIBODIES OR BIOTINYLATED PEPTIDES CORRESPONDING ...
; NUMBER OF SEQUENCES: 453
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE, P.C.
; STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR
; CITY: Arlington
; STATE: VA
; COUNTRY: USA
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/723.425A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

; NAME: SADOFF, B.J.
; REGISTRATION NUMBER: 36,663
; REFERENCE/DOCKET NUMBER: 1487-13
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-816-4000
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-723-425A-107

Query Match 15.0%; Score 20; DB 3; Length 22;
Best Local Similarity 100.0%; Pred. No. 9.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 LGGKPAIVDPKVELYQQYDE 76
Db 2 LGGKPAIVDPKVELYQQYDE 21

RESULT 4
US-09-112-206-107
; Sequence 107, Application US/09112206
; Patent No. 6210903
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: PROCESS FOR THE DETERMINATION OF PEPTIDES
; TITLE OF INVENTION: CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT EPITOPES AND THE
; TITLE OF INVENTION: IN A PROCESS FOR DETERMINATION OF ANTIBODIES OR BIOTINYLATED
; TITLE OF INVENTION: PEPTIDES CORRESPONDING TO IMMUNOLOGICALLY IMPORTANT EPITOPES
; TITLE OF INVENTION: PROCESS FOR PREPARING THEM AND COMPOSITIONS CONTAINING THEM
; NUMBER OF SEQUENCES: 453
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA: US/09/112.206
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/146.028
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-112-206-107

Query Match 15.0%; Score 20; DB 3; Length 22;
Best Local Similarity 100.0%; Pred. No. 9.1e-13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 LGGKPAIVDPKVELYQQYDE 76
Db 2 LGGKPAIVDPKVELYQQYDE 21

RESULT 5
US-08-244-116B-2
; Sequence 2, Application US/08244116B
; Patent No. 5763159
; GENERAL INFORMATION:
; APPLICANT: Simmonds, Peter
; APPLICANT: Chan, Shiu-Wan
; APPLICANT: Yap, Peng L.

;; TITLE OF INVENTION: Hepatitis-C Virus Testing
;; NUMBER OF SEQUENCES: 53
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Bell, Seltzer, Park & Gibson, P.A.
;; STREET: 1211 East Morehead Street
;; CITY: Charlotte
;; STATE: No. 5761359th Carolina
;; COUNTRY: United States
;; ZIP: 28234
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0. Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/244,116B
;; FILING DATE: 15-JUL-1994
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/GB92/02143
;; FILING DATE: 20-NOV-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Sibley, Kenneth D.
;; REGISTRATION NUMBER: 31,665
;; REFERENCE/DOCKET NUMBER: 1749-125
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 704-377-1561
;; TELEFAX: 704-334-2014
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 19 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; HYPOTHETICAL: NO
;; FRAGMENT TYPE: Internal
;; ORIGINAL SOURCE:
;; ORGANISM: Hepatitis-C virus
US-08-244-116B-2

Query Match 14.3%; Score 19; DB 1; Length 19;
Best Local Similarity 100.0%; Pred. No. 7.5e-12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 79 ECSQAAYIEQAQVIAHQF 97
Db 1 ECSQAAYIEQAQVIAHQF 19

RESULT 6
US-08-324-977-44
; Sequence 44, Application US/08324977
; Patent No. 5747339
; GENERAL INFORMATION:
; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FUKE, Isao
; APPLICANT: MORI, Chisato
; APPLICANT: TAKAMIZAWA, Akahisa
; APPLICANT: YOSHIDA, Iwao
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
; TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Armstrong, Westerman, Hattori, McLeLland &
; STREET: 1725 K St. N.W. Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in, 1.44Mb

;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
;; SOFTWARE: ASCII
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/324,977
;; FILING DATE: 18-OCT-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 2-167466
;; FILING DATE: 25-JUN-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 2-230921
;; FILING DATE: 31-AUG-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 2-305605
;; FILING DATE: 09-NOV-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/099,706
;; FILING DATE: 30-JUL-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/769,996
;; FILING DATE: 02-OCT-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/635,451
;; FILING DATE: 28-DEC-1990
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Stevens-Smith, Theresa M.
;; REGISTRATION NUMBER: 36,281
;; REFERENCE/DOCKET NUMBER: 900703D
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 659-2930
;; TELEFAX: (202) 887-0357
;; TELEX: 440142
;; INFORMATION FOR SEQ ID NO: 44:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 247 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-324-977-44

Query Match 14.3%; Score 19; DB 1; Length 247;
Best Local Similarity 100.0%; Pred. No. 8e-11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 LTHPITKYIMACMSADLEV 24
Db 22 LTHPITKYIMACMSADLEV 40

RESULT 7
US-08-384-616-44
; Sequence 44, Application US/08384616
; Patent No. 5847101
; GENERAL INFORMATION:
; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FUKE, Isao
; APPLICANT: MORI, Chisato
; APPLICANT: TAKAMIZAWA, Akahisa
; APPLICANT: YOSHIDA, Iwao
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
; TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Armstrong, Westerman, Hattori, McLeLland &
; STREET: 1725 K St. N.W. Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in, 1.44Mb
; COMPUTER: IBM PC compatible

```
; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/384,616
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/769,996
; FILING DATE: 02-OCT-1991
; APPLICATION NUMBER: JP 2-167466
; FILING DATE: 25-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-230921
; FILING DATE: 31-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-305605
; FILING DATE: 09-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/635,451
; FILING DATE: 28-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Stevens-Smith, Theresa M.
; REGISTRATION NUMBER: 36,281
; REFERENCE/DOCKET NUMBER: 900703B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 659-2930
; TELEFAX: (202) 887-0357
; TELEX: 440142
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 247 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-384-616-44

Query Match 14.3%; Score 19; DB 2; Length 247;
Best Local Similarity 100.0%; Pred. No. 8e-11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 LTHPTIKYIMACMSADLEV 24
Db 22 LTHPTIKYIMACMSADLEV 40

RESULT 8
US-08-904-686A-44
; Sequence 44, Application US/08904686A
; Patent No. 5998130
; GENERAL INFORMATION:
; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FUKU, Isao
; APPLICANT: MORI, Chisato
; APPLICANT: TAKAMIZAWA, Akahisa
; APPLICANT: YOSHIDA, Iwao
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
; TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Armstrong, Westernman, Hattori, McLeLland &
; STREET: 1725 K St. N.W. Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in, 1.44Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/904,686A
```

```
; FILING DATE: 01-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/324,977
; FILING DATE: 18-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-167466
; FILING DATE: 25-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-230921
; FILING DATE: 31-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-305605
; FILING DATE: 09-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/099,706
; FILING DATE: 30-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/769,996
; FILING DATE: 02-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/635,451
; ATTORNEY/AGENT INFORMATION:
; NAME: McLeLland, Le-Nhung
; REGISTRATION NUMBER: 31,541
; REFERENCE/DOCKET NUMBER: 900703G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 659-2930
; TELEFAX: (202) 887-0357
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 247 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-904-686A-44

Query Match 14.3%; Score 19; DB 2; Length 247;
Best Local Similarity 100.0%; Pred. No. 8e-11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 LTHPTIKYIMACMSADLEV 24
Db 22 LTHPTIKYIMACMSADLEV 40

RESULT 9
US-09-315-850-44
; Sequence 44, Application US/09315850
; Patent No. 6217872
; GENERAL INFORMATION:
; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FUKU, Isao
; APPLICANT: MORI, Chisato
; APPLICANT: TAKAMIZAWA, Akahisa
; APPLICANT: YOSHIDA, Iwao
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
; TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Armstrong, Westernman, Hattori, McLeLland &
; STREET: 1725 K St. N.W. Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in, 1.44Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/315,850-44
```

APPLICATION NUMBER: US/09/315,850
FILING DATE: 23-AUG-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: DE FRANCESCO-1
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 631 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-700-356-1

Query Match 14.3%; Score 19; DB 3; Length 247;
Best Local Similarity 100.0%; Pred. No. 8e-11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 LTHPITKYIMACMSADLEV 24
|||||
Db 22 LTHPITKYIMACMSADLEV 40

RESULT 10
US-08-700-356-1
Sequence 1, Application US/08700356
Patent No. 5739002
GENERAL INFORMATION:
APPLICANT: DE FRANCESCO, Raffaele
APPLICANT: FAILLA, Cristina
APPLICANT: TOMEI, Licia
TITLE OF INVENTION: METHOD FOR REPRODUCING IN VITRO THE
TELEPHONE: (202) 887-0357
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/700,356
FILING DATE: 23-AUG-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: DE FRANCESCO-1
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 631 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-700-356-1

Query Match 14.3%; Score 19; DB 1; Length 631;
Best Local Similarity 100.0%; Pred. No. 1.9e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 LTHPITKYIMACMSADLEV 24
|||||
Db 611 LTHPITKYIMACMSADLEV 629

RESULT 11
US-08-936-865-1
Sequence 1, Application US/08936865
Patent No. 5861297
GENERAL INFORMATION:
APPLICANT: Sardana, Vinod V
APPLICANT: Blue, Jeffrey T
TITLE OF INVENTION: DETERGENT-FREE HEPATITIS C PROTEASE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: MERCK & CO., INC.
STREET: P.O. Box 2000, 126 E. Lincoln Ave.
CITY: Rahway
STATE: NJ
COUNTRY: US
ZIP: 07065-0907
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,865
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Avler, Sylvia A
REGISTRATION NUMBER: 36,436
REFERENCE/DOCKET NUMBER: 19691
TELEPHONE: 908-594-4909
TELEFAX: 908-594-4720
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 631 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: Hepatitis C Virus
STRAIN: NS3 Serine Protease Domain

```

; Patent No. 6211338
; GENERAL INFORMATION:
; APPLICANT: Malcolim, Bruce
; APPLICANT: Taremi, Shahrar S.
; APPLICANT: Weber, Patricia
; APPLICANT: Yao, Nanhua
; TITLE OF INVENTION: Covalent Complexes of Hepatitis C Virus
; TITLE OF INVENTION: NS3 Protease and NS4A Cofactor Peptide
; NUMBER OF SEQUENCES: 123
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-Plough Corp.
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07030
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Power Macintosh
; OPERATING SYSTEM: 8.0.1
; SOFTWARE: Microsoft Word 6.0.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/198,723A
; FILING DATE: 24 NOV 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McLaughlin, Jaye P.
; REGISTRATION NUMBER: 41,211
; REFERENCE/DOCKET NUMBER: JB0800
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)298-5056
; TELEFAX: (908)298-5388
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 646 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-198-723A-60

Query Match 14.3%; Score 19; DB 3; Length 646;
Best Local Similarity 100.08; Pred. No. 1.9e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0

QY 6 LTHPTTKYIMACMSADLEV 24
   |||||
DDB 626 LTHPTTKYIMACMSADLEV 644

RESULT 14
US-09-198-723A-63
; Sequence 63, Application US/09198723A
; Patent No. 6211338
; GENERAL INFORMATION:
; APPLICANT: Malcolim, Bruce
; APPLICANT: Taremi, Shahrar S.
; APPLICANT: Weber, Patricia
; APPLICANT: Yao, Nanhua
; TITLE OF INVENTION: Covalent Complexes of Hepatitis C Virus
; TITLE OF INVENTION: NS3 Protease and NS4A Cofactor Peptide
; NUMBER OF SEQUENCES: 123
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-Plough Corp.
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07030
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Power Macintosh
; OPERATING SYSTEM: 8.0.1
; SOFTWARE: Microsoft Word 6.0.1

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 11:23:35 ; Search time 20 seconds
(without alignments)
909.600 Million cell updates/sec

Title: US-09-638-693-36

Perfect score: 133

Sequence: 1 QNEICLTHPTIKYIMACMSA.....VIEPIVTTNQKLEAFWHKH 133

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 510680 seqs, 136781880 residues

Word size : 0

Total number of hits satisfying chosen parameters: 510680

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	133	100.0	133	11	US-09-899-046-36
2	133	100.0	133	11	US-09-878-281-36
3	98	73.7	133	11	US-09-899-046-38
4	98	73.7	133	11	US-09-899-046-40
5	98	73.7	133	11	US-09-878-281-38
6	98	73.7	133	11	US-09-878-281-40
7	98	73.7	209	11	US-09-899-046-223
8	98	73.7	209	11	US-09-878-281-223
9	52	39.1	133	11	US-09-899-046-32
10	52	39.1	133	11	US-09-899-046-34
11	52	39.1	133	11	US-09-878-281-32
12	52	39.1	133	11	US-09-878-281-34
13	41	30.8	829	10	US-09-881-239-5
14	41	30.8	1099	10	US-09-881-654-4
15	20	15.0	20	11	US-09-899-046-97

Sequence 97, Appl
Sequence 30, Appl
Sequence 30, Appl
Sequence 270, App
Sequence 270, App
Sequence 198, App
Sequence 200, App
Sequence 198, App
Sequence 200, App
Sequence 4, Appl
Sequence 11, Appl
Sequence 18, Appl
Sequence 2, Appl
Sequence 9, Appl
Sequence 16, Appl
Sequence 40, Appl
Sequence 99, Appl
Sequence 99, Appl
Sequence 42, Appl
Sequence 3, Appl
Sequence 3, Appl
Sequence 98, Appl
Sequence 100, App
Sequence 98, Appl
Sequence 100, App
Sequence 91, Appl
Sequence 7, Appl
Sequence 30, Appl
Sequence 7, Appl

US-09-899-046-36
; Sequence 36, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF INVENTIONS: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 133 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-899-046-36

Query Match 100.0%; Score 133; DB 11; Length 133;
Best Local Similarity 100.0%; Pred. No. 2.8e-121;
Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QNEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSGVCVIVGHIELGK 60
Db 1 QNEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSGVCVIVGHIELGK 60
QY 61 PAIVPDKVLYQOYDEMEECQAAPVIEQAQVTAHQFKGLVGLLQQRATQQAQVIEPIVT 120

ALIGNMENTS

RESULT 1

US-09-899-046-36
; Sequence 36, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF INVENTIONS: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 133 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-899-046-36

Query Match 100.0%; Score 133; DB 11; Length 133;
Best Local Similarity 100.0%; Pred. No. 2.8e-121;
Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QNEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSGVCVIVGHIELGK 60
Db 1 QNEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSGVCVIVGHIELGK 60
QY 61 PAIVPDKVLYQOYDEMEECQAAPVIEQAQVTAHQFKGLVGLLQQRATQQAQVIEPIVT 120

```
Db 61 PAIVDPKVELYQOYDEMECSQAAPYIEQAQVIAHOFKGVGLLQQRATQQAQVIEPIVT 120
Qy 121 TNWQLEAFWHKH 133
    |||||
Db 121 TNWQLEAFWHKH 133

RESULT 2
US-09-878-281-36
; Sequence 36, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 133 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-878-281-36

Query Match 100.0%; Score 133; DB 11; Length 133;
Best Local Similarity 100.0%; Pred. No. 2.8e-121;
Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNEICLTHPTIKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
Db 1 QNEICLTHPTIKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
Qy 61 PAIVDPKVELYQOYDEMECSQAAPYIEQAQVIAHOFKGVGLLQQRATQQAQVIEPIVT 120
Db 61 PAIVDPKVELYQOYDEMECSQAAPYIEQAQVIAHOFKGVGLLQQRATQQAQVIEPIVT 120
Qy 121 TNWQLEAFWHKH 133
    |||||
Db 121 TNWQLEAFWHKH 133

RESULT 3
US-09-899-046-40
; Sequence 38, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 133 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-899-046-40

Query Match 100.0%; Score 98; DB 11; Length 133;
Best Local Similarity 100.0%; Pred. No. 2.4e-87;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNEICLTHPTIKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
Db 1 QNEICLTHPTIKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
Qy 61 PAIVDPKVELYQOYDEMECSQAAPYIEQAQVIAHOFKGVGLLQQRATQQAQVIEPIVT 120
Db 61 PAIVDPKVELYQOYDEMECSQAAPYIEQAQVIAHOFKGVGLLQQRATQQAQVIEPIVT 120
Qy 121 TNWQLEAFWHKH 133
    |||||
Db 121 TNWQLEAFWHKH 133

RESULT 4
US-09-899-046-40
; Sequence 40, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 133 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-899-046-40

Query Match 73.7%; Score 98; DB 11; Length 133;
Best Local Similarity 100.0%; Pred. No. 2.4e-87;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QNEICLTHPTIKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
Db 1 QNEICLTHPTIKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
Qy 61 PAIVDPKVELYQOYDEMECSQAAPYIEQAQVIAHOFK 98
Db 61 PAIVDPKVELYQOYDEMECSQAAPYIEQAQVIAHOFK 98

RESULT 5
US-09-878-281-38
; Sequence 38, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```


RESULT 7
US-09-899-046-223
; Sequence 223, Application US/098999046

Qy	1	QNEICLTHPITKYIMACSADLEVTSTWLLGGVLAALAAAYCLSVGCVVVGHIELGK	60
Db	77	QNEICLTHPITKYIMACSADLEVTSTWLLGGVLAALAAAYCLSVGCVVVGHIELGK	136

Best Local Similarity 100.0%; Pred. No. 9.1e-43;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 TRYINACMSADLEVTSTWVLGGVLAALAAAYCLSVGCVIVGHIELGKKA 62
Db 11 TRYINACMSADLEVTSTWVLGGVLAALAAAYCLSVGCVIVGHIELGKKA 62

RESULT 13

US-09-881-239-5
; Sequence 5, Application US/09881239
; Publication No. US20020192639A1
; GENERAL INFORMATION:
; APPLICANT: CHIEN, David Y.
; APPLICANT: ARCANDEL, Phillip
; APPLICANT: TANDESKE, Laura
; APPLICANT: GEORGE-NASCIEMENTO, Carlos
; APPLICANT: COIT, Doris
; APPLICANT: MEDINA-SELBY, Angelica
; TITLE OF INVENTION: HCV ANTIGEN/ANTIBODY COMBINATION ASSAY
; FILE REFERENCE: 2302-16073 / PP16073.003
; CURRENT APPLICATION NUMBER: US/09/881,239
; CURRENT FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MEFA 12
US-09-881-239-5

Query Match 30.8%; Score 41; DB 10; Length 829;
Best Local Similarity 100.0%; Pred. No. 2.2e-31;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GKGPAIVDPKVELYQYDEMERCSSQAAPYIEQAQVIAHQFK 98
Db 444 GKGPAIVDPKVELYQYDEMERCSSQAAPYIEQAQVIAHQFK 484

RESULT 14

US-09-881-654-4
; Sequence 4, Application US/09881654
; Patent No. US20020146685A1
; GENERAL INFORMATION:
; APPLICANT: CHIEN, David Y.
; APPLICANT: ARCANDEL, Phillip
; APPLICANT: TANDESKE, Laura
; APPLICANT: GEORGE-NASCIEMENTO, Carlos
; APPLICANT: COIT, Doris
; APPLICANT: MEDINA-SELBY, Angelica
; TITLE OF INVENTION: IMMUNOASSAYS FOR ANTI-HCV ANTIBODIES
; FILE REFERENCE: 2302-17039 / PP17039.002
; CURRENT APPLICATION NUMBER: US/09/881,654
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: 60/212,082
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 60/280,811
; PRIOR FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: 60/280,867
; PRIOR FILING DATE: 2001-04-02
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1099
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MEFA 7.1
US-09-881-654-4

Query Match 30.8%; Score 41; DB 10; Length 1099;
Best Local Similarity 100.0%; Pred. No. 2.8e-31;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 GKGPAIVDPKVELYQYDEMERCSSQAAPYIEQAQVIAHQFK 98
Db 748 GKGPAIVDPKVELYQYDEMERCSSQAAPYIEQAQVIAHQFK 788

RESULT 15

US-09-899-046-97
; Sequence 97, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 97:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: BR36
; POSITION IN GENOME:
; MAP POSITION: Positions 1688 to 1707 of HCV type 3
US-09-899-046-97

Query Match 15.0%; Score 20; DB 11; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.7e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 LGGKPAIVDPKVELYQYQYDE 76
Db 1 LGGKPAIVDPKVELYQYQYDE 20

Search completed: August 29, 2003, 11:25:57
Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 29, 2003, 12:21:40 ; Search time 3076 Seconds

(without alignments)
1768.849 Million cell updates/sec

Title: US-09-638-693-36

Perfect score: 133

Sequence: 1 QNEICLTHPTIKYIMACMSA.....VIEPIVTTNWQKLEAFWHKH 133

Scoring table:
OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Word size: 1

Total number of hits satisfying chosen parameters: 5773148

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n model -DEV=xlh
-Q/cgn2_1/USPTO.spool/US09638693/runat_25082003_173340_22262/app_query.fasta_1.327
-DB=GenEmbl -QFMT=fastap -SUFFIX=oli.rge -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09638693 @CCN_1_1_3508 @runat_25082003_173340_22262 -NCPU=6 -ICPU=3
-NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*			
1:	gb.ba.*	22	52
2:	gb.htg.*	23	52
3:	gb.in.*	24	52
4:	gb.om.*	25	52
5:	gb.ov.*	26	52
6:	gb.pat.*	27	52
7:	gb.ph.*	28	52
8:	gb.pl.*	29	52
9:	gb.pr.*	30	52
10:	gb.ro.*	31	52
11:	gb.sts.*	32	52
12:	gb.sy.*	33	52
13:	gb.un.*	34	52
14:	gb.vi.*	35	52
15:	em.ba.*	36	48
16:	em.fun.*	37	43
17:	em.in.*	38	41
18:	em.mu.*	39	41
19:	em.om.*	40	35
20:	em.or.*	41	35
21:	em.ov.*	42	35
22:	em.pat.*	43	35
23:	em.ph.*	44	35
24:	em.pl.*	45	35
25:	em.ro.*		
26:	em.sts.*		
27:	em.un.*		
28:	em.un.*		

29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htg_mus.*
34: em.htg_pln.*
35: em.htg_rod.*
36: em.htg_mam.*
37: em.htg_vrt.*
38: em.sy.*
39: em.htgo_hum.*
40: em.htgo_mus.*
41: em.htgo_other.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	133	100.0	401	6	A40635 Sequence 35
2	133	100.0	401	6	AX031613 Sequence
3	133	100.0	401	6	AX031883 Sequence
4	133	100.0	401	6	AX032153 Sequence
5	133	100.0	401	6	BD172141 New sequ
6	98	73.7	401	6	A40637 Sequence 37
7	98	73.7	401	6	A40639 Sequence 39
8	98	73.7	401	6	AX031615 Sequence
9	98	73.7	401	6	AX031617 Sequence
10	98	73.7	401	6	AX031885 Sequence
11	98	73.7	401	6	AX031887 Sequence
12	98	73.7	401	6	AX032155 Sequence
13	98	73.7	401	6	AX032157 Sequence
14	98	73.7	401	6	BD172142 New sequ
15	98	73.7	401	6	BD172143 New sequ
16	98	73.7	629	6	A40822 Sequence 22
17	98	73.7	629	6	AX031800 Sequence
18	98	73.7	629	6	AX032070 Sequence
19	98	73.7	629	6	AX032340 Sequence
20	98	73.7	629	6	BD172226 New sequ
21	98	73.7	629	14	HPCNS4EE D14600 Hepatitis C
22	52	39.1	401	6	A40631 Sequence 31
23	52	39.1	401	6	A40633 Sequence 33
24	52	39.1	401	6	AX031609 Sequence
25	52	39.1	401	6	AX031611 Sequence
26	52	39.1	401	6	AX031879 Sequence
27	52	39.1	401	6	AX031881 Sequence
28	52	39.1	401	6	AX032149 Sequence
29	52	39.1	401	6	AX032151 Sequence
30	52	39.1	401	6	BD172139 New sequ
31	52	39.1	401	6	BD172140 New sequ
32	52	39.1	1000	14	HPCNS4G D14602 Hepatitis C
33	52	39.1	1000	14	HCVNS34 X79819 Hepatitis C
34	52	39.1	9390	14	HVCVENS1 X76918 Hepatitis C
35	48	36.1	367	6	AR011609 Sequence
36	48	36.1	9425	14	AF046866 Hepatitis
37	43	32.3	9456	14	HPCEGS D17763 Hepatitis C
38	41	30.8	2499	6	AX395311 Sequence
39	41	30.8	2520	12	AF127762 Synthetic
40	41	30.8	3297	6	AX454820 Sequence
41	35	26.3	210	14	AF007504 Hepatitis
42	35	26.3	264	14	AF007503 Hepatitis
43	35	26.3	270	14	AF007509 Hepatitis
44	35	26.3	270	14	AF007511 Hepatitis
45	35	26.3	270	14	AF007513 Hepatitis

ALIGNMENTS

A40635
LOCUS A40635 401 bp DNA linear PAT 05-MAR-1997
DEFINITION Sequence 35 from Patent WO9425601.
ACCESSION A40635
VERSION A40635.1 GI:2296670
KEYWORDS
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 401)
AUTHORS Maertens,G. and Stuyver,L.
TITLE NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE AS
THERAPEUTIC AND DIAGNOSTIC AGENTS
JOURNAL Patent: WO 9425601-A 35 10-NOV-1994;
INNOGENETICS NV (BE)
COMMENT Other publication CA 2139100 941110
Other publication AU 6722294 941121
Other publication CN 1108030 950906
Other publication FI 946066 941223
Other publication NO 944967 941221
Other publication JP 7508423T 950921.
FEATURES
source Location/Qualifiers
1..401
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/clone="BR36-20-164"
3..401
/note="unnamed protein product; Protein sequence is in
conflict with the conceptual translation"
/codon_start=1
/protein_id="CAA02504.1"
/db_xref="GI:2296671"
/translation="QNEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCL
SVGCVTVGHIELGKPAIVDPDKVLYQQYDEMECSQAAPYIEQAQVIAHQFKGKVL
GLQRATQQQAIVPIVITNNQKLEAFWHKH"
BASE COUNT 110 a 100 c 101 g 90 t
ORIGIN
Alignment Scores:
Pred. No.: 3e-127 Length: 401
Score: 133.00 Matches: 133
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-638-693-36 (1-133) x A40635 (1-401)
Qy 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
Db 3 CAAAAATGAATCTGCTTGACACACCCCATCACAAATACATCATGGCATGTCAGCT 62
Qy 21 AspleuGluValThrSerThrTrpValLeuLeuGlyGlyValLeuAlaAlaLeuAla 40
Db 63 GATCTGGAAGTAACACACAGACCTGGGTTTGGTGGAGGGGTCCTCGCGGCCCTAGCG 122
Qy 41 AlatyrcysLeuSerValGlyCysValIleValGlyHisIleGluLeuGlyGlyLys 60
Db 123 GCCTACTGCTGTGTCAGTCGGTTGTTGTTGATTGTGGTTCATATCGAGCTGGGGGCAAG 182
Qy 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGlyCys 80
Db 183 CCGGCAATCGTTCAGACAAAGAGGTGTTGTATCAACAATACGATGAGTGAAGAGTGC 242
Qy 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100
Db 243 TCACAGCTGCCCATATATCGAACAGCTCAGGTATAGCTCACCAGTTCACAGGGA 302
Qy 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValThr 120
Db 303 GTCCTTGGATTGTCAGCGAGCCACCAACAACAGCTGTCTATTGAGCCCATAGTA 362
Qy 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133

121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
Db 363 ACCAACTGGCAAAAGCTTGAGCCCTTTTGGCAAGCAT 401
RESULT 2
LOCUS AX031613 401 bp DNA linear PAT 20-SEP-2000
DEFINITION Sequence 35 from Patent EP1004670.
ACCESSION AX031613
VERSION AX031613.1 GI:10278850
KEYWORDS
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1
AUTHORS Maertens,G. and Stuyver,L.
TITLE Sequences of hepatitis c virus genotypes and their use as
therapeutic and diagnostic agents
JOURNAL Patent: EP 1004670-A 35 31-MAY-2000;
INNOGENETICS NV (BE)
FEATURES
source Location/Qualifiers
1..401
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/clone="BR36-20-164"
3..401
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC09682.1"
/db_xref="GI:10278851"
/translation="QNEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCL
SVGCVTVGHIELGKPAIVDPDKVLYQQYDEMECSQAAPYIEQAQVIAHQFKGKVL
GLQRATQQQAIVPIVITNNQKLEAFWHKH"
BASE COUNT 110 a 100 c 101 g 90 t
ORIGIN
Alignment Scores:
Pred. No.: 3e-127 Length: 401
Score: 133.00 Matches: 133
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-638-693-36 (1-133) x AX031613 (1-401)
Qy 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
Db 3 CAAAAATGAATCTGCTTGACACACCCCATCACAAATACATCATGGCATGTCAGCT 62
Qy 21 AspleuGluValThrSerThrTrpValLeuLeuGlyGlyValLeuAlaAlaLeuAla 40
Db 63 GATCTGGAAGTAACACACAGACCTGGGTTTGGTGGAGGGGTCCTCGCGGCCCTAGCG 122
Qy 41 AlatyrcysLeuSerValGlyCysValIleValGlyHisIleGluLeuGlyGlyLys 60
Db 123 GCCTACTGCTGTGTCAGTCGGTTGTTGTTGATTGTGGTTCATATCGAGCTGGGGGCAAG 182
Qy 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGlyCys 80
Db 183 CCGGCAATCGTTCAGACAAAGAGGTGTTGTATCAACAATACGATGAGTGAAGAGTGC 242
Qy 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100
Db 243 TCACAGCTGCCCATATATCGAACAGCTCAGGTATAGCTCACCAGTTCACAGGGA 302
Qy 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValThr 120
Db 303 GTCCTTGGATTGTCAGCGAGCCACCAACAACAGCTGTCTATTGAGCCCATAGTA 362
Qy 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133

```
Db 363 ACCAACTGGCAAAAGCTTGAGGCCTTTTGGCACAAGCAT 401
RESULT 3
AX031883
LOCUS AX031883
DEFINITION Sequence 35 from Patent EP0984068.
ACCESSION AX031883
VERSION AX031883.1 GI:10279033
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Maertens,G. and Stuyver,L.
TITLE Sequences of hepatitis c virus genotypes and their use as
JOURNAL therapeutic and diagnostic agents
INNOGENETICS NV (BE) Patent: EP 0984068-A 35 08-MAR-2000;
FEATURES
source
1. .401
Location/Qualifiers
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/clone="BR36-20-164"
<3. .>401
CDS
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC09759.1"
/db_xref="GI:10279034"
/translation="QNEICLTHPTIKYIMACMSADLEVTSTWVLLGGVLAALAAAYCL
SVGCVIVGHIELGKPAIVPKDEVLYQYDEMECSQAAPYIEQAQVIAHQFKGKVL
GLQRAQQAQVIEPIVTNNWOKLEAFWKKH"
BASE COUNT 110 a 100 c 101 g 90 t
ORIGIN
Alignment Scores:
Pred. No.: 3e-127 Length: 401
Score: 133.00 Matches: 133
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-638-693-36 (1-133) x AX031883 (1-401)
QY 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
Db 3 CAAAATGAATCTGCTTGACACACCCCATCACAATAATACATCATGCGATGTCAGCT 62
QY 21 AspLeuGluValThrThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLeuAla 40
Db 63 GATCTGGAAGTAACCAACAGCACCCTGGGTTTGTGGAGGGTCTCTCGCGCCCTAGCG 122
QY 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyGlyLys 60
Db 123 GCCTACTGCTTGCAGTCAGTGTGTGTGATTGTGGTGCATATCGAGCTGGGGGCAAG 182
QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGluCys 80
Db 183 CCGGCATCGTCCAGACAAAGAGGTGTGTATCAACAATACGATGAGTGAAGAGTGC 242
QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPhelysGlyLys 100
Db 243 TCACAAGCTGCCCATATATCGAACCAAGCTCAGTAAATAGCTCACCAGTTCAAGGGA 302
QY 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValThr 120
Db 303 GTCCCTGGATTGCTGCAGCGAGCCACCAACAACAGCTGTCTATTGAGCCCATAGTA 362
QY 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
Db 363 ACCAACTGGCAAAAGCTTGAGGCCTTTTGGCACAAGCAT 401
RESULT 5
BD172141
```

```
RESULT 4
AX032153
LOCUS AX032153
DEFINITION Sequence 35 from Patent EP0984067.
ACCESSION AX032153
VERSION AX032153.1 GI:10279216
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Maertens,G. and Stuyver,L.
TITLE Sequences of hepatitis c virus genotypes and their use as
JOURNAL therapeutic and diagnostic agents
INNOGENETICS NV (BE) Patent: EP 0984067-A 35 08-MAR-2000;
FEATURES
source
1. .401
Location/Qualifiers
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/clone="BR36-20-164"
<3. .>401
CDS
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC09835.1"
/db_xref="GI:10279217"
/translation="QNEICLTHPTIKYIMACMSADLEVTSTWVLLGGVLAALAAAYCL
SVGCVIVGHIELGKPAIVPKDEVLYQYDEMECSQAAPYIEQAQVIAHQFKGKVL
GLQRAQQAQVIEPIVTNNWOKLEAFWKKH"
BASE COUNT 110 a 100 c 101 g 90 t
ORIGIN
Alignment Scores:
Pred. No.: 3e-127 Length: 401
Score: 133.00 Matches: 133
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-638-693-36 (1-133) x AX032153 (1-401)
QY 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
Db 3 CAAAATGAATCTGCTTGACACACCCCATCACAATAATACATCATGCGATGTCAGCT 62
QY 21 AspLeuGluValThrThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLeuAla 40
Db 63 GATCTGGAAGTAACCAACAGCACCCTGGGTTTGTGGAGGGTCTCTCGCGCCCTAGCG 122
QY 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyGlyLys 60
Db 123 GCCTACTGCTTGCAGTCAGTGTGTGTGATTGTGGTGCATATCGAGCTGGGGGCAAG 182
QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGluCys 80
Db 183 CCGGCATCGTCCAGACAAAGAGGTGTGTATCAACAATACGATGAGTGAAGAGTGC 242
QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPhelysGlyLys 100
Db 243 TCACAAGCTGCCCATATATCGAACCAAGCTCAGTAAATAGCTCACCAGTTCAAGGGA 302
QY 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValThr 120
Db 303 GTCCCTGGATTGCTGCAGCGAGCCACCAACAACAGCTGTCTATTGAGCCCATAGTA 362
QY 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
Db 363 ACCAACTGGCAAAAGCTTGAGGCCTTTTGGCACAAGCAT 401
RESULT 5
BD172141
```

LOCUS	BD172141	401 bp	DNA	linear	PAT 18-FEB-2003
DEFINITION	New sequences of hepatitis C virus genotypes for diagnosis, prophylaxis and therapy.				
ACCESSION	BD172141				
VERSION	BD172141.1	GI:284113439			
KEYWORDS	JP 2002233389-A/18.				
SOURCE	unidentified				
ORGANISM	unclassified.				
REFERENCE	1 (bases 1 to 401)				
AUTHORS	Maertens,G. and Stuyver,L.				
TITLE	New sequences of hepatitis C virus genotypes for diagnosis, prophylaxis and therapy				
JOURNAL	Patent: JP 2002233389-A 18 20-AUG-2002;				
COMMENT	NV INNOGENETICS SA				
	OS Unidentified				
	PN JP 2002233389-A/18				
	PD 20-AUG-2002				
	PF 21-NOV-2001 JP 2001356707				
	PR 27-APR-1993 EP 93401099.2, 05-AUG-1993 EP 93402019.9 PI				
	GEERT MAERTENS, LIEVEN STUYVER				
	PC C12N15/09, A61K35/76, A61K38/00, A61K39/00, A61K39/395, A61K48/00,				
	PC A61P31/20, C07K14/18, C07K16/10, C12Q1/68, G01N33/53, G01N33/566,				
	PC G01N33/576, C12N15/00, A61K37/02				
	CC Strandedness: Single;				
	CC Topology: Linear;				
	CC New sequences of hepatitis C virus genotypes for diagnosis, and therapy				
	CC prophylaxis				
FEATURES	Location/Qualifiers				
source	1..401				
	/organism="unidentified"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:32644"				
BASE COUNT	110 a 100 c 101 g 90 t				
ORIGIN					
Alignment Scores:					
Pred. No.:	3e-127	Length:	401		
Score:	133.00	Matches:	133		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		
Query Match:	100.00%	Indels:	0		
DB:	6	Gaps:	0		
US-09-638-693-36 (1-133) x BD172141 (1-401)					
QY	1	GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla	20		
DB	3	CAAAATGAATCTGCTTGACACACCCATCAAAATACATCATGCGATGTCAGCT	62		
QY	21	AspLeuGluValThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLaLeuAla	40		
DB	63	GATCTGGAAGTAACCAACACGACCTGGGTTTGTCTGGAGGGTCTCGCGCCCTAGCG	122		
QY	41	AlaTyrCysLeuSerValGlyCysValIleValGlyHisIleGluLeuGlyGlyLys	60		
DB	123	GCCTACTGCTTGTACGTGGTGTGTGTGATTGTGGTGCATATCGAGCTGGGGGCAAG	182		
QY	61	ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluCys	80		
DB	183	CCGGCAATCGTTCCAGACAAAGAGTGTGTATCAAAATACGATGAGTGAAGAGTGC	242		
QY	81	SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys	100		
DB	243	TCACAAGCTGCCCATATATCGACACAGCTCAGGTATAGCTACCAAGTTCAGGGANA	302		
QY	101	ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValThr	120		
LOCUS	BD172141	401 bp	DNA	linear	PAT 18-FEB-2003
DEFINITION	New sequences of hepatitis C virus genotypes for diagnosis, prophylaxis and therapy.				
ACCESSION	BD172141				
VERSION	BD172141.1	GI:284113439			
KEYWORDS	JP 2002233389-A/18.				
SOURCE	unidentified				
ORGANISM	unclassified.				
REFERENCE	1 (bases 1 to 401)				
AUTHORS	Maertens,G. and Stuyver,L.				
TITLE	New sequences of hepatitis C virus genotypes for diagnosis, prophylaxis and therapy				
JOURNAL	Patent: JP 2002233389-A 18 20-AUG-2002;				
COMMENT	NV INNOGENETICS SA				
	OS Unidentified				
	PN JP 2002233389-A/18				
	PD 20-AUG-2002				
	PF 21-NOV-2001 JP 2001356707				
	PR 27-APR-1993 EP 93401099.2, 05-AUG-1993 EP 93402019.9 PI				
	GEERT MAERTENS, LIEVEN STUYVER				
	PC C12N15/09, A61K35/76, A61K38/00, A61K39/00, A61K39/395, A61K48/00,				
	PC A61P31/20, C07K14/18, C07K16/10, C12Q1/68, G01N33/53, G01N33/566,				
	PC G01N33/576, C12N15/00, A61K37/02				
	CC Strandedness: Single;				
	CC Topology: Linear;				
	CC New sequences of hepatitis C virus genotypes for diagnosis, and therapy				
	CC prophylaxis				
FEATURES	Location/Qualifiers				
source	1..401				
	/organism="unidentified"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:32644"				
BASE COUNT	110 a 100 c 101 g 90 t				
ORIGIN					
Alignment Scores:					
Pred. No.:					


```

|||||
Db 243 TCACAAGCTGCCCATATATCGAACAAAGCTCAGGTGATAGCTCACCAGTTCAA -GGAAAA 301
|||||
Qy 100 sValleuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValTh 120
|||||
Db 302 AGTCCTTGATTCGTCGACGCGAGCCACCAACAAGCTGTCAATTGAGCCCATAGTAAC 361
|||||
Qy 120 rThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
|||||
Db 362 TACCACCTGGCAAAAGCTTGAGCGCCTTTTGGCACAAGCAT 401
|||||
RESULT 7
A40639
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
OTHER
Other publication CA 2139100 941110
Other publication AU 6722294 941121
Other publication CN 1108030 950906
Other publication FI 946066 941223
Other publication NO 944967 941221
Other publication JP 7508423T 950921.
FEATURES
source
1..401
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/clone="BR36-20-165"
3..401
/note="unnamed protein product; Protein sequence is in
conflict with the conceptual translation"
/codon_start=1
/protein_id="CAA02506.1"
/db_xref="GI:2296675"
/translation="QNEICLTHPTIKYIMACMSADLEVTSTWVLLGGVLAALAAAYCL
SYGCVVIVGHIELGKPAIVDPKEVLYQYDEMEECSQAAPYIEQAQVIAHOFKEKVL
GLLQRTQQQAVIEPIVTNNQLEAFWHKH"
BASE COUNT 111 a 100 c 100 g 90 t
ORIGIN
Alignment Scores:
Pred. No.: 2,49e-91 Length: 401
Score: 98.00 Matches: 132
Percent Similarity: 98.51% Conservative: 0
Best Local Similarity: 98.51% Mismatches: 1
Query Match: 73.68% Indels: 2
DB: 6 Gaps: 0
US-09-638-693-36 (1-133) x A40639 (1-401)
Qy 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
|||||
Db 3 CAATATGAATCTGCTTGACACACCCCATCACAAATACATCATGCGATGTCAGCT 62
|||||
Qy 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyGlyValLeuAlaAlaLeuAla 40
|||||
Db 63 GATCTGGAAGTAACCAACAGCACCTGGGTTTTCCTTGGAGGGGTCTTCGCGCCCTAGCG 122
|||||
Qy 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyLys 60
|||||
Db 123 GCCTACTGCTTCAGTCGGTGTGTGTGATTGTGGTTCATATCGAGCTGGGGGCAAG 182
|||||

```

```

Qy 61 ProAlaIleValProAspLysGluValLeuTyrGlnInTyrAspGluMetGluGluCys 80
|||||
Db 183 CCGGCAATCGTTCACAGACAAAGAGGTGTGTATCAACAATACATGATGAGATGAAGAGTGC 242
|||||
Qy 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGly-Ly 100
|||||
Db 243 TCACAAGCTGCCCATATATCGAACAAAGCTCAGGTAATAGTCCACCAGTTCAA -GGAAAA 301
|||||
Qy 100 sValleuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValTh 120
|||||
Db 302 AGTCCTTGATTCGTCGACGCGAGCCACCAACAAGCTGTCAATTGAGCCCATAGTAAC 361
|||||
Qy 120 rThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
|||||
Db 362 TACCACCTGGCAAAAGCTTGAGCGCCTTTTGGCACAAGCAT 401
|||||
RESULT 8
AX031615
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1..401
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/clone="BR36-20-166"
<3..>401
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC09683.1"
/db_xref="GI:10278853"
/translation="QNEICLTHPTIKYIMACMSADLEVTSTWVLLGGVLAALAAAYCL
SYGCVVIVGHIELGKPAIVDPKEVLYQYDEMEECSQAAPYIEQAQVIAHOFKEKVL
GLLQRTQQQAVIEPIVTNNQLEAFWHKH"
BASE COUNT 110 a 100 c 101 g 90 t
ORIGIN
Alignment Scores:
Pred. No.: 2,49e-91 Length: 401
Score: 98.00 Matches: 132
Percent Similarity: 98.51% Conservative: 0
Best Local Similarity: 98.51% Mismatches: 1
Query Match: 73.68% Indels: 2
DB: 6 Gaps: 0
US-09-638-693-36 (1-133) x AX031615 (1-401)
Qy 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
|||||
Db 3 CAATATGAATCTGCTTGACACACCCCATCACAAATACATCATGCGATGTCAGCT 62
|||||
Qy 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyGlyValLeuAlaAlaLeuAla 40
|||||
Db 63 GATCTGGAAGTAACCAACAGCACCTGGGTTTTCCTTGGAGGGGTCTTCGCGCCCTAGCG 122
|||||
Qy 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyLys 60
|||||
Db 123 GCCTACTGCTTCAGTCGGTGTGTGTGATTGTGGTTCATATCGAGCTGGGGGCAAG 182
|||||
Qy 61 ProAlaIleValProAspLysGluValLeuTyrGlnInTyrAspGluMetGluGluCys 80
|||||

```


Db 243 TCACAAAGCTGCCCATATATCGAACAAAGCTCAGGTGATAGCTCACCAGTTCAA-GGAAAA 301
 QY 100 sValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValTh 120
 Db 302 AGTCCCTGGATTGCTCGAGCGAGCCACCCCAACAACAGCTGTCTATTGAGCCCATAGTAAC 361
 QY 120 rThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
 Db 362 TACCAACTGGCAAAAGCTTGAGGCCCTTTGGCACAGCAT 401

RESULT 11
 AX031887
 LOCUS
 DEFINITION
 AX031887
 ACCESSION
 AX031887.1 GI:10279037
 KEYWORDS
 SOURCE
 ORGANISM
 Maertens, G. and Stuyver, L.
 Sequences of hepatitis C virus genotypes and their use as
 therapeutic and diagnostic agents
 Patent: EP 0984068-A 39 08-MAR-2000;
 INNOGENETICS NV (BE)

FEATURES
 source
 1..401
 /organism="unidentified"
 /mol_type="genomic DNA"
 /db_xref="taxon:32644"
 /clone="BR36-20-165"
 <3..>401
 /codon_start=1
 /protein_id="CAC09761.1"
 /db_xref="GI:10279038"
 /translation="QNEICLTHPTIKYIMACMSADLEVTTSWLLGGVLAALAAVCL
 SVGCVVVGHIELGKPAIVDPKEVLYQQYDEMEECSQAAPYIEQAQVIAHQFKEVCL
 GLLQRATQQQAVIEPIVTNNWQLEAFWHKH"
 BASE COUNT 111 a 100 c 100 g 90 t
 ORIGIN

Alignment Scores:
 Pred. No.: 2,49e-91 Length: 401
 Score: 98.00 Matches: 132
 Percent Similarity: 98.51% Conservative: 0
 Best Local Similarity: 98.51% Mismatches: 1
 Query Match: 73.68% Indels: 2
 DB: 6 Gaps: 0

US-09-638-693-36 (1-133) x AX031887 (1-401)

QY 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
 Db 3 CAAATGAATCTGCTTGACACACCCCATCACAAATACATCATGCGATGCTAGCT 62
 QY 21 AspLeuGluValThrThrSerThrTrpValLeuLeuGlyValLeuAlaLeuAla 40
 Db 63 GATCTGGAAGTACCAACACACACCTGGTGTTCCTGGAGGGTCTCGCGGCCCTAGCG 122
 QY 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyLys 60
 Db 123 GCCTACTGCTGTGACGTGGTGTGTGTGATTGTGGTGTATATCGAGCTGGGGGCAAG 182
 QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluCys 80
 Db 183 CCGCAATCTCTCCAGCAAGAGGTGTTGTATCAACAATACCATGAGATGGAAGAGTGC 242
 QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGly-Ly 100
 Db 243 TCACAAAGCTGCCCATATATCGACACAGCTCAGGTATAGCTCACCAGTTCAA-GGAAAA 301

QY 100 sValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValTh 120
 Db 302 AGTCCCTGGATTGCTCGAGCGAGCCACCCCAACAACAGCTGTCTATTGAGCCCATAGTAAC 361
 QY 120 rThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
 Db 362 TACCAACTGGCAAAAGCTTGAGGCCCTTTGGCACAGCAT 401

RESULT 12
 AX032155
 LOCUS
 DEFINITION
 AX032155
 ACCESSION
 AX032155.1 GI:10279218
 KEYWORDS
 SOURCE
 ORGANISM
 Maertens, G. and Stuyver, L.
 Sequences of hepatitis C virus genotypes and their use as
 therapeutic and diagnostic agents
 Patent: EP 0984067-A 37 08-MAR-2000;
 INNOGENETICS NV (BE)

FEATURES
 source
 1..401
 /organism="unidentified"
 /mol_type="genomic DNA"
 /db_xref="taxon:32644"
 /clone="BR36-20-166"
 <3..>401
 /codon_start=1
 /protein_id="CAC09836.1"
 /db_xref="GI:10279219"
 /translation="QNEICLTHPTIKYIMACMSADLEVTTSWLLGGVLAALAAVCL
 SVGCVVVGHIELGKPAIVDPKEVLYQQYDEMEECSQAAPYIEQAQVIAHQFKEVCL
 GLLQRATQQQAVIEPIVTNNWQLEAFWHKH"
 BASE COUNT 110 a 100 c 101 g 90 t
 ORIGIN

Alignment Scores:
 Pred. No.: 2,49e-91 Length: 401
 Score: 98.00 Matches: 132
 Percent Similarity: 98.51% Conservative: 0
 Best Local Similarity: 98.51% Mismatches: 1
 Query Match: 73.68% Indels: 2
 DB: 6 Gaps: 0

US-09-638-693-36 (1-133) x AX032155 (1-401)

QY 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
 Db 3 CAAATGAATCTGCTTGACACACCCCATCACAAATACATCATGCGATGCTAGCT 62
 QY 21 AspLeuGluValThrThrSerThrTrpValLeuLeuGlyValLeuAlaLeuAla 40
 Db 63 GATCTGGAAGTACCAACACACACCTGGTGTTCCTGGAGGGTCTCGCGGCCCTAGCG 122
 QY 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyLys 60
 Db 123 GCCTACTGCTGTGACGTGGTGTGTGTGATTGTGGTGTATATCGAGCTGGGGGCAAG 182
 QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluCys 80
 Db 183 CCGCAATCTCTCCAGCAAGAGGTGTTGTATCAACAATACCATGAGATGGAAGAGTGC 242
 QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGly-Ly 100
 Db 243 TCACAAAGCTGCCCATATATCGACACAGCTCAGGTATAGCTCACCAGTTCAA-GGAAAA 301
 QY 100 sValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValTh 120

D	b		302	AGTCCTTGGATVGTGTCAGGCAGGCCACCCCAACAACAAAGCTGTGTCATTGAGCCCCATAGTAAC	361
O	y		120	rThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis	133
D	b		362	TACCAACTGCGAAGCTTGAGGCTTTTGGCACAGCAT	401
R	E	S	U	L	T
RESULT	13				
AX032157					
LOCUS					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
FEATURES					
source					
CDS					
BASE COUNT					
ORIGIN					
Alignment Scores:					
Pred. No.:					
Score:					
Percent Similarity:					
Best Local Similarity:					
Query Match:					
DB:					
US-09-638-693-36	(1-133)	x	AX032157	(1-401)	
Oy	1	GlaAsnclurleCysLeuThrHisProilleThrIstYrileMetaLaCysMetSerAla	20		
D	b	3	CAAATGAATCTGCTTGACACACCCCACATCACAAAATACATCATGCGATGTCATGTCCAGCT	62	
Oy	21	AspLeuGluValThrThrSerThrTrpValLeuGlyGlyValLeuAlaLeuAla	40		
D	b	63	GATCTGGAAGTAACCACACACCTGGGTTTTCTTGGAGGGGTCTCGCGGCCCTAGCG	122	
Oy	41	AlatyrCysLeuSerValGlyCysValVallevalcylHislleGluLeuGlyLys	60		
D	b	123	GCCTACTGCVTGTGTCAGTCGGTGTGTGTGTTGTTGATGTGGGTTCATATCGAGCTGGGGGCAAG	182	
Oy	61	ProAlatleValProAspLysGluValLeuTyrglnInTyrraspGluMetGluGluCys	80		
D	b	183	CCGGCAATCGTTCAGACAAAGAGGTGTGTATCAACAATACCATGAGATGGAAGAGTGC	242	
Oy	81	SerGlnAlaAlaProTrylleGluGlnAlaGlnValIleAlaHisGlnPheLysGly-Ly	100		
D	b	243	TCACAAGCTGCCCATATATCGAACAGCTCAGGTAACTAGCTCACCAGTTCAA-GGAAAA	301	
Oy	100	svalLeuglyLeuLeuGlnArgalathrGlnGlnGlnAlavalilleGluProfileValTh	120		
D	b	302	AGTCCTTGGATTGCTGAGCAGCAGCCACCAACAACAGCTGTCTATTGAGCCCATAGTAAC	361	

[illegible]

QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGly-Ly 100
 |||||
 Db 243 TCACAAAGCTGCCCATATATCAACAAGCTCAGGTGATAGCTCACCAGTTCAA-GGAAAA 301
 |||||
 QY 100 sValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValTh 120
 |||||
 Db 302 AGTCCTTGGATTGCTGCAGCAGGCCACCAACAAGCTGTCTTGGAGCCCATAGTAAC 361
 |||||
 QY 120 rThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
 |||||
 Db 362 TACCAACTGGCAAAAGCTTGAGGCCTTTTGGCACAAGCAT 401
 |||||

RESULT 15

BD172143
 LOCUS 401 bp DNA linear PAT 18-FEB-2003
 DEFINITION New sequences of hepatitis C virus genotypes for diagnosis,
 prophylaxis and therapy.

ACCESSION

BD172143

VERSION

BD172143.1 GI:28413441

KEYWORDS

JP 2002233389-A/20.

SOURCE

unidentified

ORGANISM

unclassified.

REFERENCE

1 (bases 1 to 401)

AUTHORS

Maertens,G. and Stuyver,L.

TITLE

New sequences of hepatitis C virus genotypes for diagnosis,

JOURNAL

prophylaxis and therapy

Patent: JP 2002233389-A 20-AUG-2002;

COMMENT

NV INNOGENETICS SA

OS Unidentified

PN JP 2002233389-A/20

PD 20-AUG-2002

PF 21-NOV-2001 JP 2001356707

PR 27-APR-1993 EP 93401099.2, 05-AUG-1993 EP 93402019.9 PI

GEERT MAERTENS,LIEVEN STUYVER

PC C12N15/09,A61K35/76,A61K38/00,A61K39/00,A61K39/395,A61K39/53,

A61K48/00,

PC A61P31/20,C07K14/18,C07K16/10,C12Q1/68,G01N33/53,G01N33/53, PC

G01N33/566

PC G01N33/576,C12N15/00,A61K37/02

CC Strandedness: Single;

CC Topology: Linear;

CC New sequences of hepatitis C virus genotypes for diagnosis,

prophylaxis

CC and therapy

EH Key

FT CDS

Location/Qualifiers

1. .401

Location/Qualifiers

3. .401.

/organism="unidentified"

/mol_type="genomic DNA"

/db_xref="taxon:32644"

111 a 100 c 100 g 90 t

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 2,49e-91 Length: 401
 Score: 98.00 Matches: 132
 Percent Similarity: 98.51% Conservative: 0
 Best Local Similarity: 98.51% Mismatches: 1
 Query Match: 73.68% Indels: 2
 DB: 6 Gaps: 0

US-09-638-693-36 (1-133) x BD172143 (1-401)

QY 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20

|||||

Db 3 CAAATGAATCTGCTTGACACACCCCATCACAATAATCATATGCGCATGTCAGCT 62

|||||

QY 21 AspLeuGluValThrThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLeuAla 40

|||||

Db 63 GATCTGGAAGTAACCAACACGACCTGGGTTTTGCTTGAGGGGTCTCGCGGCCCTAGCG 122

QY 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyGlyLys 60
 |||||
 Db 123 GCCTACTGCTTGTCTAGTCGTTGTTGTTGTTGGTTCATATCGAGCTGGGGGCGAAG 182
 |||||
 QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGlyCys 80
 |||||
 Db 183 CCGGCAATCGTTCCAGACAAAGAGGTGTGTATCAACAATACGATGAGATGGAAGAGTGC 242
 |||||
 QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGly-Ly 100
 |||||
 Db 243 TCACAAAGCTGCCCATATATCGAACAAGCTCAGGTAATAGCTCACCAGTTCAA-GGAAAA 301
 |||||
 QY 100 sValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValTh 120
 |||||
 Db 302 AGTCCTTGGATTGCTGCAGCAGGCCACCAACAAGCTGTCTTGGAGCCCATAGTAAC 361
 |||||
 QY 120 rThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
 |||||
 Db 362 TACCAACTGGCAAAAGCTTGAGGCCTTTTGGCACAAGCAT 401
 |||||

Search completed: August 29, 2003, 13:52:42

Job time : 3078 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 29, 2003, 11:30:20 ; Search time 240 Seconds
(without alignments)
1495.939 Million cell updates/sec

Title: US-09-638-693-36
Perfect score: 133
Sequence: 1 QNEICLTHPTIKYIMACMSA.....VIEPIVTTNWKLEAFWHKH 133

Scoring table:
OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues
Word size: 1

Total number of hits satisfying chosen parameters: 5103490

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-O=/cn2.1/USPTO.spool/US09638693/runat_25082003_173340_22254/app_query.fasta_1.327
-DB=N_Geneseq_19Jun03 -QFMT=fastp -SUFFIX=oli.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09638693.ecgn_1.1.312.0runat_25082003_173340_22254 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DRV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N_Geneseq_19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	133	100.0	401	15	AAQ78040 Hepatitis C virus
2	98	73.7	401	15	AAQ78041 Hepatitis C virus
3	98	73.7	401	15	AAQ78042 Hepatitis C virus
4	98	73.7	629	15	AAQ78125 HCV sequence used
5	88	66.2	367	14	AAQ43111 HCV NS4 coding reg
6	88	66.2	367	14	AAQ43106 HCV NS4 coding reg
7	55	41.4	355	14	AAQ43108 HCV NS4 coding reg
8	52	39.1	365	14	AAQ43107 HCV NS4 coding reg
9	52	39.1	401	15	AAQ78038 Hepatitis C virus
10	52	39.1	401	15	AAQ78039 Hepatitis C virus
11	48	36.1	353	14	AAQ43109 HCV NS4 coding reg
12	48	36.1	363	14	AAQ43110 HCV NS4 coding reg
13	41	30.8	2499	24	AAQ29796 Multiple epitope f
14	41	30.8	3297	24	ABK15345 HCV multiple epit
15	41	30.8	3297	25	ABK14411 DNA encoding HCV m
16	30	22.6	9444	17	AAQ78037 HCV antigen. Synt
17	19	14.3	287	15	AAQ78037 HCV antigen. Synt
18	19	14.3	582	15	AAQ62690 HCV gene 10. Hepa
19	19	14.3	586	13	AAQ26990 DNA associated wit
20	19	14.3	588	14	AAV05564 HK10. Hepatitis C
21	19	14.3	669	13	AAQ27012 HCV sequence used
22	19	14.3	1443	15	AAQ78126 Hepatitis C virus
23	19	14.3	1470	15	AAQ78115 HCV NS4A-NS3 compl
24	19	14.3	1998	20	AAQ80353 HCV NS4A-NS3 compl
25	19	14.3	1998	20	AAQ80354 HCV NS4A-NS3 compl
26	19	14.3	1998	20	AAQ80355 HCV NS4A-NS3 compl
27	19	14.3	1998	20	AAQ80356 HCV NS4A-NS3 compl
28	19	14.3	1998	20	AAQ80357 HCV NS4A-NS3 compl
29	19	14.3	1998	20	AAQ80358 HCV NS4A-NS3 compl
30	19	14.3	1998	20	AAQ80359 HCV NS4A-NS3 compl
31	19	14.3	1998	20	AAQ80352 HCV NS4A-NS3 compl
32	19	14.3	2007	20	AAQ80369 HCV NS4A-NS3 compl
33	19	14.3	2007	20	AAQ80370 HCV NS4A-NS3 compl
34	19	14.3	2007	20	AAQ80366 HCV NS4A-NS3 compl
35	19	14.3	2007	20	AAQ80367 HCV NS4A-NS3 compl
36	19	14.3	2007	20	AAQ80368 HCV NS4A-NS3 compl
37	19	14.3	2013	20	AAQ80360 HCV NS4A-NS3 compl
38	19	14.3	2016	20	AAQ80361 HCV NS4A-NS3 compl
39	19	14.3	6973	14	AAV05577 DNA associated wit
40	19	14.3	9413	16	AAQ78040 Partial HCV non-st
41	19	14.3	9413	16	AAQ81559 Hepatitis C virus
42	19	14.3	9413	16	AAQ80498 DNA encoding HCV p
43	19	14.3	9413	24	AAQ25517 Hepatitis C virus
44	19	14.3	9413	25	AAQ49655 Hepatitis C virus
45	19	14.3	9413	25	AAQ53723 Hepatitis C virus

ALIGNMENTS

RESULT 1
AAQ78040
ID AAQ78040 standard; cdna; 401 BP.
XX
AC AAQ78040;
XX
DT 25-MAR-2003 (updated)
DT 01-AUG-1995 (first entry)
XX
Hepatitis C virus NS3/NS4 region.
XX
Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
KW classification; immunisation; prophylaxis; serotyping; ss.
XX
Hepatitis C virus type 3.
OS
XX
FH Key Location/Qualifiers

```
FT CDS 3..401
FT /*tag= a
FT /product= NS3/NS4 polypeptide.
XX W09425601-A2.
XX 10-NOV-1994.
XX 27-APR-1994; 94WO-EP01323.
XX 27-APR-1993; 93EP-0401099.
XX 05-AUG-1993; 93EP-0402019.
XX (INNO-) INNOGENETICS NV SA.
XX Maertens G, Stuyver L;
XX WPI: 1994-358277/44.
XX P-PSDB; AAR63288.
XX New polynucleotide sequences from hepatitis C virus - and related
XX vectors, polypeptide(s) and antibodies, useful for immunisation,
XX treatment, diagnosis and typing of HCV isolates
XX Claim 2; Page 124-125; 404pp; English.
XX Compositions comprising at least 5, and pref. 8 or more contiguous
XX nucleotides selected from an HCV type 3 genomic sequence, more
XX particularly (i) the region spanning positions 417-957 of the
XX Core/E1 region of HCV subtype 3a; (ii) the region spanning positions
XX 4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
XX positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
XX region spanning positions 8023-8235 of the NS5 region of the BR36
XX subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
XX isolate, may be used as primers to amplify nucleic acid from an
XX isolate belonging to a specific genotype, or as a probe for specific
XX detection/classification of nucleic acid. Polypeptides encoded by
XX the nucleotides in such compositions may be used for immunisation
XX against HCV, for the detection of antibodies directed against HCV
XX and for serotyping. This sequence corresponds to the NS3/NS4
XX region of HCV subtype 3a and is taken from a clone designated
XX BR36-20-164.
XX (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 401 BP; 110 A; 100 C; 101 G; 90 T; 0 other;

Alignment Scores:
Pred. No.: 5,76e-126 Length: 401
Score: 133.00 Matches: 133
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-09-638-693-36 (1-133) x AAQ78040 (1-401)
. Qy 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetaIleCysMetSerAla 20
Db 3 CAATAAGAAATCTGCTTGACACACCCCATCAAAATACATCATGGCATGTCAGCT 62
Qy 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyValLeuAlaLeuAla 40
Db 63 GATCTGGAAGTAACACACACACACCTGGGTTTCTGGAGGGGTCCTCGGCCCTAGCG 122
Qy 41 AlaTyrCysLeuSerValGlyCysValIleValGlyHisIleGluLeuGlyGlyLys 60
Db 123 GCCTACTCTGCTGTCAGTCGGTGTGTGATTGTCGTCATATCGAGCTGGGGGCAAG 182
Qy 61 ProAlaIleValProAspLysGluValIleuTyrGlnGlnTyrAspGluMetGluGluCys 80
Db 183 CCGCAATCGTTCAGACAAAGAGGTGTTGATCAACAATACCATGAGATGGAAGATGC 242
Qy 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100
```

```
Db 243 TCACAAGCTGCCCATATATCGAACAGCTCAGTAAATAGCTCACCAGTTCAAGGAAAA 302
Qy 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValThr 120
Db 303 GTCCTTGATGTCGACGAGCCACCACCAACAAAGCTGTCATTTAGGCCCATAGTA 362
Qy 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
Db 363 ACCAACTGGCAAAAGCTTGAGCGCTTTTGGCACAAGCAT 401
RESULT 2
AAQ78041
ID AAQ78041 standard; cDNA; 401 BP.
XX AC AAQ78041;
XX 25-MAR-2003 (updated)
XX 01-AUG-1995 (first entry)
XX Hepatitis C virus NS3/NS4 region.
XX Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
XX classification; immunisation; prophylaxis; serotyping; ss.
XX Hepatitis C virus type 3.
XX Key Location/Qualifiers
XX CDS 3..401
XX /*tag= a
XX /product= NS3/NS4 polypeptide.
XX W09425601-A2.
XX 10-NOV-1994.
XX 27-APR-1994; 94WO-EP01323.
XX 27-APR-1993; 93EP-0401099.
XX 05-AUG-1993; 93EP-0402019.
XX (INNO-) INNOGENETICS NV SA.
XX Maertens G, Stuyver L;
XX WPI: 1994-358277/44.
XX P-PSDB; AAR63288.
XX New polynucleotide sequences from hepatitis C virus - and related
XX vectors, polypeptide(s) and antibodies, useful for immunisation,
XX treatment, diagnosis and typing of HCV isolates
XX Claim 2; Page 126; 404pp; English.
XX Compositions comprising at least 5, and pref. 8 or more contiguous
XX nucleotides selected from an HCV type 3 genomic sequence, more
XX particularly (i) the region spanning positions 417-957 of the
XX Core/E1 region of HCV subtype 3a; (ii) the region spanning positions
XX 4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
XX positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
XX region spanning positions 8023-8235 of the NS5 region of the BR36
XX subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
XX isolate, may be used as primers to amplify nucleic acid from an
XX isolate belonging to a specific genotype, or as a probe for specific
XX detection/classification of nucleic acid. Polypeptides encoded by
XX the nucleotides in such compositions may be used for immunisation
XX against HCV, for the detection of antibodies directed against HCV
XX and for serotyping. This sequence corresponds to the NS3/NS4
XX region of HCV subtype 3a and is taken from a clone designated
XX BR36-20-166.
XX (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 401 BP; 110 A; 100 C; 101 G; 90 T; 0 other;
```



```
Alignment Scores:
Pred. No.: 2,05e-90 Length: 401
Score: 98.00 Matches: 132
Percent Similarity: 98.51% Conservatives: 0
Best Local Similarity: 98.51% Mismatches: 1
Query Match: 73.68% Indels: 2
DB: 15 Gaps: 0

US-09-638-693-36 (1-133) x AAQ78041 (1-401)
QY 1 GlnAsnGluLeuCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
DB 3 CAAATGAATCTGCTTGACACCCCATACAAAAATACATCATGTCATGTCAGCT 62
QY 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLeuAla 40
DB 63 GATCTGAAGTACACACACACACCTGCTGCTTGGAGGGTCTCTCGGGGCCCTAGCG 122
QY 41 AlaTyrCysLeuSerValGlyCysValValIleValIleGlyHisIleGluLeuGlyLys 60
DB 123 GCCTACTGCTTGTCAAGTCGGTGTGTGATTGTGGTTCATATCGAGCTGGGGGCAAG 182
QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGluCys 80
DB 183 CCGGCAATCGTTCACAGACAAAGAGGTGTGTATCAACAATACATGATGGAAGAGTGC 242
QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGly-Ly 100
DB 243 TCACAAGCTGCCCATATATCGAACCAAGCTCAGCTAATAGCTCACCAGTTCAA-GGAAAA 301
QY 100 sValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValTh 120
DB 302 AGTCCTTGGATTGCTGCAGCGAGCCACCCAAACAAGCTGTCTATTGAGCCCATAGTAAC 361
QY 120 rThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
DB 362 TACCAACTGGCAAAAGCTTGAGGCCTTTTGGCACAAGCAT 401

RESULT 4
AAQ78125
ID AAQ78125 standard; cDNA; 629 BP.
XX
XX AAQ78125;
XX
```

```
WPI; 1994-358277/44.
P-PSDB; AAR63290.
New polynucleotide sequences from hepatitis C virus - and related
vectors, polypeptide(s) and antibodies, useful for immunisation,
treatment, diagnosis and typing of HCV isolates
Claim 2; Page 128; 404pp; English.
Compositions comprising at least 5, and pref. 8 or more contiguous
nucleotides selected from an HCV type 3 genomic sequence, more
particularly (i) the region spanning positions 417-957 of the
Core/E1 region of HCV subtype 3a; (ii) the region spanning positions
4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
region spanning positions 8023-8235 of the NS5 region of the BR36
subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
sequence, may be used as primers to amplify nucleic acid from an
isolate belonging to a specific genotype, or as a probe for specific
detection/classification of nucleic acid. Polypeptides encoded by
the nucleotides in such compositions may be used for immunisation
against HCV, for the detection of antibodies directed against HCV
and for serotyping. This sequence corresponds to the NS3/NS4
region of HCV subtype 3a and is taken from a clone designated
BR36-20-165.
(Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 401 BP; 111 A; 100 C; 100 G; 90 T; 0 other;

Alignment Scores:
Pred. No.: 2,05e-90 Length: 401
Score: 98.00 Matches: 132
Percent Similarity: 98.51% Conservatives: 0
Best Local Similarity: 98.51% Mismatches: 1
Query Match: 73.68% Indels: 2
DB: 15 Gaps: 0

US-09-638-693-36 (1-133) x AAQ78042 (1-401)
QY 1 GlnAsnGluLeuCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
DB 3 CAAATGAATCTGCTTGACACCCCATACAAAAATACATCATGTCATGTCAGCT 62
QY 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLeuAla 40
DB 63 GATCTGAAGTACACACACACACCTGCTGCTTGGAGGGTCTCTCGGGGCCCTAGCG 122
QY 41 AlaTyrCysLeuSerValGlyCysValValIleValIleGlyHisIleGluLeuGlyLys 60
DB 123 GCCTACTGCTTGTCAAGTCGGTGTGTGATTGTGGTTCATATCGAGCTGGGGGCAAG 182
QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGluCys 80
DB 183 CCGGCAATCGTTCACAGACAAAGAGGTGTGTATCAACAATACATGATGGAAGAGTGC 242
QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGly-Ly 100
DB 243 TCACAAGCTGCCCATATATCGAACCAAGCTCAGGTGATAGCTCACCAGTTCAA-GGAAAA 301
QY 100 sValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValTh 120
DB 302 AGTCCTTGGATTGCTGCAGCGAGCCACCCAAACAAGCTGTCTATTGAGCCCATAGTAAC 361
QY 120 rThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
DB 362 TACCAACTGGCAAAAGCTTGAGGCCTTTTGGCACAAGCAT 401

AAQ78042 standard; cDNA; 401 BP.
XX
AC AAQ78042;
XX
XX 25-MAR-2003 (updated)
DT 01-AUG-1995 (first entry)
XX
XX Hepatitis C virus NS3/NS4 region.
XX
XX Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
KW classification; immunisation; prophylaxis; serotyping; ss.
XX
OS Hepatitis C virus type 3.
XX
XX Key Location/Qualifiers
FH 3..401
FT CDS /tag= a
FT /product= NS3/NS4 polypeptide.
XX
XX WO9425601-A2.
PN
XX
XX 10-NOV-1994.
PD
XX
XX 27-APR-1994; 94WO-EP01323.
PF
XX
XX 27-APR-1993; 93EP-0401099.
PR
XX 05-AUG-1993; 93EP-0402019.
XX
XX (INNO-) INNOGENETICS NV SA.
PA
XX
XX Maertens G, Stuyver L;
PI
XX
```


CC type 3 being phylogenetically different to type 1 and 2. The same
 CC degree differentiation was noted in the NS-5 (see AAR37923-26) and core
 CC region between type 3 and type 1 sequences.
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 367 BP; 95 A; 93 C; 101 G; 78 U; 0 other;

Alignment Scores:

Pred. No.: 2,72e-80 Length: 367
 Score: 88.00 Matches: 88
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 66.17% Indels: 0
 DB: 14 Gaps: 0

US-09-638-693-36 (1-133) x AAQ43111 (1-367)

QY 11 ThrLysTyrIleMetAlaCysMetSerAlaAspLeuGluValThrSerThrTrpVal 30
 DB 14 ACAAAUAUCAUGCAUGCAUGCAGCUGAUGAAGUAACCAACAGCAGCAGCAGGUG 73
 QY 31 LeuLeuGlyGlyValLeuAlaLeuAlaTyrCysLeuSerValGlyCysValVal 50
 DB 74 UUGCUUGGAGGGGUCUGCGCGCCGACGCGCCUACUGCUGCAGCUGCGUGUG 133
 QY 51 IleValGlyHisIleGluLeuGlyGlyLysProAlaIleValProAspLysGluValLeu 70
 DB 134 AUUGUGGUCAUAUUGCUGCGGGGCGAGCGCGCAUUGUCCAGACAAGAGGUGUG 193
 QY 71 TyrGlnGlnTyrAspGluMetGluGlyCysSerGlnAlaAlaProTyrIleGluGlnAla 90
 DB 194 UAUCAACAUAUGAUGAGAGGAGGAGGUGCUGCAAGCUGCGCCCAUAUAUCCAAGCU 253
 QY 91 GlnValIleAlaHisGlnPheLys 98
 DB 254 CAGGUGAUGCCACCAGUCCAAG 277

RESULT 6

AAQ43106
 ID AAQ43106 standard; DNA; 367 BP.

XX AAQ43106;

DT 25-MAR-2003 (updated)
 DT 23-SEP-1993 (first entry)

XX HCV NS4 coding region from donor T0040.

XX Non-coding region; hepatitis C virus; blood donor; type 2; type 1;
 KW HCV; NS-5; phylogeny; differentiation; NS-3; core region; type 3; ss.

OS Hepatitis C virus.

XX Key Location/Qualifiers
 FT CDS 2..367
 FT /*tag= a

XX WO9310239-A2.

XX 27-MAY-1993.

XX 20-NOV-1992; 92WO-GB02143.

XX 21-NOV-1991; 91GB-0024696.

XX 24-JUN-1992; 92GB-0013362.

XX (COMM-) COMMON SERVICES AGENCY.

XX Chan S, Simmonds P, Yap PL;

XX WPI; 1993-182554/22.

XX P-PSDB; AAR37932.

XX

PT DNA encoding antigenic peptide(s) of new types of hepatitis C
 PT virus - for diagnosing and treating HCV infection, screening
 PT blood samples and identifying different HCV types
 XX Disclosure; Fig 9a; 120pp; English.

CC The sequences given in AAQ43106-111 show bases 4911-5277 of the NS4
 CC region of hepatitis C virus-3 (HCV-3) samples from 5 blood donors
 CC and a consensus sequence. Analysis of this and other regions of the
 CC HCV genome revealed the existence of three distinct groups of HCV.
 CC Analysis of the region encompassing -255 to -62 of the 5' non coding
 CC region (see AAQ43058-75) showed a difference of 9-14% in the nucleotide
 CC sequences between the three groups. Two of the groups identified were
 CC similar to those of HCV variants termed type 1 and 2, whilst the third
 CC appeared to represent a novel type of virus. Comparison of the NS3
 CC region (see AAR37927-30) showed a high degree of sequence diversity with
 CC type 3 being phylogenetically different to type 1 and 2. The same
 CC degree differentiation was noted in the NS-5 (see AAR37923-26) and core
 CC region between type 3 and type 1 sequences.
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 367 BP; 96 A; 92 C; 101 G; 78 U; 0 other;

Alignment Scores:

Pred. No.: 2,72e-80 Length: 367
 Score: 88.00 Matches: 88
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 66.17% Indels: 0
 DB: 14 Gaps: 0

US-09-638-693-36 (1-133) x AAQ43106 (1-367)

QY 11 ThrLysTyrIleMetAlaCysMetSerAlaAspLeuGluValThrSerThrTrpVal 30
 DB 14 ACAAAUAUCAUGCAUGCAUGCAGCUGAUGAAGUAACCAACAGCAGCAGCAGGUG 73
 QY 31 LeuLeuGlyGlyValLeuAlaLeuAlaTyrCysLeuSerValGlyCysValVal 50
 DB 74 UUGCUUGGAGGGGUCUGCGCGCCGACGCGCCUACUGCUGCAGCUGCGUGUG 133
 QY 51 IleValGlyHisIleGluLeuGlyGlyLysProAlaIleValProAspLysGluValLeu 70
 DB 134 AUUGUGGUCAUAUUGCUGCGGGGCGAAGCGCGCAUUGUCCAGACAAGAGGUGUG 193
 QY 71 TyrGlnGlnTyrAspGluMetGluGlyCysSerGlnAlaAlaProTyrIleGluGlnAla 90
 DB 194 UAUCAACAUAUGAUGAGAGGAGGAGGUGCUGCAAGCUGCGCCCAUAUAUCCAAGCU 253
 QY 91 GlnValIleAlaHisGlnPheLys 98
 DB 254 CAGGUGAUGCCACCAGUCCAAG 277

RESULT 7

AAQ43108
 ID AAQ43108 standard; DNA; 355 BP.

XX AAQ43108;

XX 25-MAR-2003 (updated)

DT 23-SEP-1993 (first entry)

XX HCV NS4 coding region from donor T0036.

XX Non-coding region; hepatitis C virus; blood donor; type 2; type 1;
 KW HCV; NS-5; phylogeny; differentiation; NS-3; core region; type 3; ss.

OS Hepatitis C virus.

XX Key Location/Qualifiers

FT CDS 3..355

XX /*tag= a

PS Claim 2; Page 122-123; 404pp; English.

XX Compositions comprising at least 5, and pref. 8 or more contiguous
CC nucleotides selected from an HCV type 3 genomic sequence, more
-CC particularly (i) the region spanning positions 417-957 of the
CC Core/E1 region of HCV subtype 3a; (ii) the region spanning positions
CC 4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
CC positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
CC region spanning positions 8023-8235 of the NS5 region of the BR36
CC subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
CC sequence, may be used as primers to amplify nucleic acid from an
CC isolate belonging to a specific genotype, or as a probe for specific
CC detection/classification of nucleic acid. Polypeptides encoded by
CC the nucleotides in such compositions may be used for immunisation
CC against HCV, for the detection of antibodies directed against HCV
CC and for serotyping. This sequence corresponds to the NS3/NS4
CC region of HCV subtype 3a and is taken from a clone designated
CC HD10-1-3.
XX (Updated on 25-MAR-2003 to correct PN field.)
XX Sequence 401 BP; 106 A; 108 C; 104 G; 83 T; 0 other;
SQ

Alignment Scores:
Pred. No.: 1.09e-43 Length: 401
Score: 52.00 Matches: 87
Percent Similarity: 97.75% Conservatives: 0
Best Local Similarity: 97.75% Mismatches: 1
Query Match: 39.10% Indels: 2
DB: 15 Gaps: 0

US-09-638-693-36 (1-133) x AAQ78039 (1-401)

QY 11 ThrLysTyrIleMetAlaCysMetSerAlaAspLeuGluValThr-Thr-SerThrtPVal 30
Db 33 ACAAAATACATTATGCATCGCATGCTGGAGTAACCCAGACCCTGGGTG 92

QY 31 LeuLeuGlyCysValLeuAlaAlaLeuAlaTyrcysLeuSerValcylcysValVal 50
Db 93 TTGCTTTGGAGGGTGCTCTCGGCCCTACTGCTGTCAGTCGGCTGCTGTA 152

QY 51 IleValcylHisIleGluLeuGlyLysProAlaIle-ValProAspLysGluValLe 70
Db 153 ATCGTGGGTTCATATCGATCGCTGGGGGCAAGCCGGCACT-CGTTCCAGACAAGGAGGTGTT 211

QY 70 uTyrGlnGlnTyrAspGluMetGluCysSerGlnAlaAlaProTyrIleGluGlnAl 90
Db 212 GTATCAACAGTAGCATGATGAGAGTGCTCGCAAGCCGCCCATATCATCATGAACAGC 271

QY 90 aGlnValIleAlaHisGlnPhelys 98
Db 272 TCAGGTAATAGCCCCACCAGCTTCAAG 296

RESULT 11
AAQ43109 standard; DNA; 353 BP.
ID AAQ43109 standard; DNA; 353 BP.
XX AAQ43109;
AC AC
XX AC
XX AC
DT 25-MAR-2003 (updated)
DT 23-SEP-1993 (first entry)
XX HCV NS4 coding region from donor T0026.
DE DE
XX DE
KW Non-coding region; hepatitis C virus; blood donor; type 2; type 1;
KW HCV; NS-5; phylogeny; differentiation; NS-3; core region; type 3; ss.
XX Hepatitis C virus.
OS OS
XX OS
XX WO9310239-A2.
PN PN
XX PN
PD 27-MAY-1993.
XX XX
PF 20-NOV-1992; 92WO-GB02143.

```
DE HCV NS4 coding region from donor T1787.
XX
KW Non-coding region; hepatitis C virus; blood donor; type 2; type 1;
KW HCV; NS-5; phylogeny; differentiation; NS-3; core region; type 3; ss.
XX
OS Hepatitis C virus.
XX
XX WO9310239-A2.
XX
XX 27-MAY-1993.
XX
XX 20-NOV-1992; 92WO-GB02143.
XX
XX 21-NOV-1991; 91GB-0024696.
XX
XX 24-JUN-1992; 92GB-0013362.
XX
XX (COMM-) COMMON SERVICES AGENCY.
XX
XX Chan S, Simmonds P, Yap RL;
XX
XX WPI: 1993-182554/22.
XX
XX P-PSDB; AAR37936.
XX
XX DNA encoding antigenic peptide(s) of new types of hepatitis C
XX virus - for diagnosing and treating HCV infection, screening
XX blood samples and identifying different HCV types
XX
XX Disclosure; Fig 9a; 120pp; English.
XX
XX The sequences given in AAQ43106-111 show bases 4911-5277 of the NS4
XX region of hepatitis C virus-3 (HCV-3) samples from 5 blood donors
XX and a consensus sequence. Analysis of this and other regions of the
XX HCV genome revealed the existence of three distinct groups of HCV.
XX CC Analysis of the region encompassing -255 to -62 of the 5' non coding
XX region (see AAQ43058-75) showed a difference of 9-14% in the nucleotide
XX sequences between the three groups. Two of the groups identified were
XX similar to those of HCV variants termed type 1 and 2, whilst the third
XX appeared to represent a novel type of virus. Comparison of the NS3
XX region (see AAR37927-30) showed a high degree of sequence diversity with
XX type 3 being phylogenetically different to type 1 and 2. The same
XX degree of differentiation was noted in the NS-5 (see AAR37923-26) and core
XX region between type 3 and type 1 sequences.
XX (Updated on 25-MAR-2003 to correct PN field.)
XX
XX SQ Sequence 363 BP; 90 A; 95 C; 102 G; 76 U; 0 other;

Alignment Scores:
Pred. No.: 1.15e-39 Length: 363
Score: 48.00 Matches: 83
Percent Similarity: 97.65% Conservative: 0
Best Local Similarity: 97.65% Mismatches: 1
Query Match: 36.09% Indels: 2
DB: 14 Gaps: 0

US-09-638-693-36 (1-133) x AAQ43110 (1-363)
QY 15 MetaLacysMetSerAlaAspLeuGluValThrThrSerThrTrpValLeuLeuGlyGly 34
DB 22 AUGGCAUGCAGUCAGUCAGUCAGUCAGUCAGUCAGUCAGUCAGUCAGUCAGUCAGUCAG 81
QY 35 ValLeuAlaAlaLeuAlaAlaTyrCysLeuSerValGlyCysValValrleValGlyHis 54
DB 82 GUUCUGCGGGCCUGACCGCCUACUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUGUCAU 141
QY 55 lleGluLeuGlyGlyLysProAlaIleValProAspLysGluValLeuTyrGlnGlnTy 74
DB 142 AUUGAGCUGGGAGGCGACCGGACU-CGUUCCACACAGAGGUGUGUUAUACACAU 200
QY 74 rAspGluMetGluGluCysSerGlnAlaAlaProTyrTleGluGlnAlaGlnValIleAl 94
DB 201 CGAUGAGUGGAGGAGUGUCGACAGCGGCCCAUAUAUCGAACAGCUCAGGUAAUAGC 260
QY 94 aHisGlnPheLys 98

Db 261 CCACCAGUUCACAG 273
|||||
RESULT 13
AAD29796
ID AAD29796 standard; DNA; 2499 BP.
XX
XX AC AAD29796;
XX
XX 17-MAY-2002 (first entry)
XX
XX Multiple epitope fusion antigen (MEFA) 12 encoding DNA.
XX
XX Hepatitis C virus; NS3/4a antigen; multiple epitope fusion antigen;
XX HCV infection; MEFA 12; ds.
XX
XX OS Unidentified.
XX
XX FH Key Location/Qualifiers
XX FT 1..2490
XX FT /*tag= a
XX FT /product= "MEFA 12 protein"
XX FT /transl_except= (pos:943..945, aa:Ile)
XX FT /transl_except= (pos:1938..1940, aa:Asp)
XX
XX PN WO200196875-A2.
XX
XX PD 20-DEC-2001.
XX
XX PF 14-JUN-2001; 2001WO-US19369.
XX
XX PR 15-JUN-2000; 2000US-212082P.
XX
XX PR 02-APR-2001; 2001US-280811P.
XX
XX PR 02-APR-2001; 2001US-280867P.
XX
XX PA (CHIR ) CHIRON CORP.
XX
XX PI Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;
XX PI Medina-Selby A;
XX
XX WPI: 2002-179522/23.
XX
XX DR P-PSDB; AAE18690.
XX
XX PT Immunoassay solid support useful for detecting hepatitis C virus
XX infection in a biological sample, comprises at least one of HCV
XX anti-core antibody and HCV NS3/4a epitope, bound to the support
XX
XX PS Disclosure; Fig 7; 87pp; English.
XX
XX The present invention relates to hepatitis C virus (HCV) core antigen
XX and NS (nonstructural) 3/4a antibody combination assay that can detect
XX both HCV antigens and antibodies present in a sample using a single
XX solid matrix as well as immunoassay solid supports for use in the assay.
XX The solid support is useful for detecting HCV infection in a biological
XX sample. The present sequence is a DNA encoding MEFA (multiple epitope
XX fusion antigen) 12 protein. This sequence is used in the exemplification
XX of the invention.
XX
XX SQ Sequence 2499 BP; 569 A; 713 C; 707 G; 510 T; 0 other;

Alignment Scores:
Pred. No.: 8.89e-32 Length: 2499
Score: 41.00 Matches: 41
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 30.83% Indels: 0
DB: 24 Gaps: 0

US-09-638-693-36 (1-133) x AAD29796 (1-2499)
QY 58 GlyGlyLysProAlaIleValProAspLysGluValLeuTyrGlnGlnTyAspGluMet 77
|||||
DB 1330 GGGGCAAGCCGGCARTCGTCCACAAAGAGGTTGTATCATACATACATGATG 1389
```

QY 78 GluGluCysSerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPhe 97
 |||||
 Db 1390 GAAGAGTCTCACAGCTGCCCATATATCGAACAGCTCAGGTAAATAGCTCACCAGTTC 1449

QY 98 Lys 98
 ||||
 Db 1450 AAG 1452

RESULT 14
 ABK15345
 ID ABK15345 standard; DNA; 3297 BP.
 XX
 AC ABK15345;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE HCV multiple epitope fusion antigen (MEFA) 7.1 gene sequence.
 XX
 KW Hepatitis C virus; HCV; NS3/4a conformational epitope; seroconversion;
 KW immunosassay solid support; multiple epitope fusion antigen; MEFA;
 KW non-structural protein; gene; ds.
 XX
 OS Hepatitis C virus.
 OS Synthetic.

Key Location/Qualifiers
 FH 1..3297
 CDS /*tag= a
 FT /partial
 FT /product= "Multiple epitope fusion antigen (MEFA) 7.1"
 FT /note= "This sequence lacks a stop codon"
 XX
 PN WO200196870-A2.
 XX
 PD 20-DEC-2001.
 XX
 PF 14-JUN-2001; 2001WO-US19156.
 XX
 PR 15-JUN-2000; 2000US-212082P.
 PR 02-APR-2001; 2001US-280811P.
 PR 02-APR-2001; 2001US-280867P.
 XX
 PA (CHIR) CHIRON CORP.
 PI Chien DY, Arcangel P, Tandeske L, George-nascimento C, Coit D;
 PI Medina-selby A;
 DR WPI: 2002-090228/12.
 DR P-PSDB: AAU76376.
 XX
 XX Immunosassay solid support, useful for detecting hepatitis C virus
 PT infection in biological sample, comprises HCV NS3/4a conformational
 PT epitope and multiple epitope fusion antigen bound to the support -
 XX
 PS Disclosure; Fig 5; 92pp; English.

CC The present invention relates to a new immunoassay solid support
 CC consisting essentially of at least one hepatitis C virus (HCV) NS3/4a
 CC conformational epitope and a multiple epitope fusion antigen (MEFA),
 CC bound to the support. The NS3/4a conformational epitope and/or
 CC MEFA reacts specifically with anti-HCV antibodies present in a biological
 CC sample from an HCV-infected individual. The immunoassay of the invention
 CC is useful for detecting hepatitis C virus infection in a biological
 CC sample. The method of the invention provides a sensitive, accurate
 CC diagnostic and prognostic tool to provide adequate patient care and to
 CC prevent transmission of HCV by blood and by blood products, or by
 CC personal contact. Use of NS3/4a conformational epitope in combination
 CC with MEFA, provides a sensitive and reliable method for detecting early
 CC HCV seroconversion. Use of MEFA has the added advantages of decreasing
 CC masking problems, improving sensitivity in detecting antibodies by
 CC allowing a greater number of epitopes on a unit surface area of
 CC substrate, and improving substrate. Detection accuracy is increased and

CC the incidence of false results is reduced because of the identification
 CC and the use of highly immunogenic HCV antigens which are present during
 CC the early stages of HCV seroconversion. The present nucleic acid sequence
 CC encodes the HCV multiple epitope fusion antigen (MEFA) 7.1 of the
 CC invention.

XX
 SQ Sequence 3297 BP; 763 A; 941 C; 915 G; 678 T; 0 other;
 Alignment Scores:
 Pred. No.: 1.15e-31 Length: 3297
 Score: 41.00 Matches: 41
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 30.83% Indels: 0
 DB: Gaps: 0

US-09-638-693-36 (1-133) x ABK15345 (1-3297)
 QY 58 GlyGlyLysProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMet 77
 |||||
 Db 2242 GGGGCAAGCGCGCAATCGTTCAGACAAAGAGGTGTGTATCAACAATAGCATGAGATG 2301

QY 78 GluGluCysSerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPhe 97
 |||||
 Db 2302 GAAGAGTCTCACAGCTGCCCATATATCGAACAGCTCAGGTAAATAGCTCACCAGTTC 2361

QY 98 Lys 98
 ||||
 Db 2362 AAG 2364

RESULT 15
 ABK14411
 ID ABK14411 standard; DNA; 3297 BP.
 XX
 AC ABK14411;
 XX
 DT 06-MAR-2003 (first entry)
 XX
 DE DNA encoding HCV multiple epitope fusion antigen 7.1 (MEFA 7.1).
 KW Immunosassay solid support; Hepatitis C virus type-1; HCV-1; HCV-2;
 KW NS3/4a conformational epitope; multiple epitope fusion antigen 7.1;
 KW MEFA 7.1; anti-HCV antibody; NS3/4a conformational antigen; HCV-3;
 KW HCV infection; Hepatitis C virus type-2; Hepatitis C virus type-3;
 KW mutant; gene; ds.
 XX
 OS Chimeric - Hepatitis C virus type 1.
 OS Chimeric - Hepatitis C virus type 2.
 OS Chimeric - Hepatitis C virus type 3.
 OS Synthetic.

Key Location/Qualifiers
 FH 1..3297
 CDS /*tag= a
 FT /partial
 FT /product= "MEFA 7.1"
 FT /note= "This sequence lacks a stop codon"
 XX
 PN US2002146685-A1.
 XX
 PD 10-OCT-2002.
 XX
 PF 14-JUN-2001; 2001US-0881654.
 XX
 PR 15-JUN-2000; 2000US-212082P.
 PR 02-APR-2001; 2001US-280811P.
 PR 02-APR-2001; 2001US-280867P.
 XX
 PA (CHIE/) CHIEN D Y.
 PA (ARCA/) ARCANDEL P.
 PA (TAND/) TANDESKE L.
 PA (GEOR/) GEORGE-NASCIMENTO C.
 PA (COIT/) COIT D.

PA (MEDI/) MEDINA-SELBY A.

XX Chien DY, Arcangel P, Tandeske L, George-Nascimento C, Coit D;
PI Medina-Selby A;

XX WPI; 2003-147573/14.
DR P-PSDB; ABG72262.

XX Immunossay solid support for detecting Hepatitis C Virus infection in
PT biological samples, comprises Hepatitis C Virus conformational epitope
PT and multiple epitope fusion antigen -

XX Claim 33; Fig 5A-5F; 45pp; English.

XX The present invention relates to immunoassays comprising Hepatitis C
CC Virus (HCV) NS3/4a conformational epitope and multiple epitope fusion
CC antigen (MEFA), bound to a solid support. The NS3/4a epitope and/or
CC the multiple epitope fusion antigen react with anti-HCV antibodies
CC present in a biological sample from an HCV-infected individual. The
CC immunoassays and methods of the invention are useful for detecting
CC HCV infection in a biological sample. The inventive immunoassay solid
CC support provides a sensitive and reliable method for detecting early
CC HCV seroconversion. The assays can detect HCV infection caused by any
CC six known genotypes of HCV. The use of the multiple epitope fusion
CC proteins decreases masking problems, improves sensitivity in detecting
CC antibodies by allowing a greater number of epitopes on a unit area
CC of substrate, and improves selectivity. The present sequence
CC encodes HCV multiple epitope fusion antigen 7.1 (MEFA 7.1), a
CC mutant HCV polypeptide derived from various regions of HCV type 1,
CC 2, or 3 (HCV-1, HCV-2, or HCV-3) polypeptide sequences.

XX Sequence 3297 BP; 763 A; 941 C; 915 G; 678 T; 0 other;

Alignment Scores:

Pred. No.:	1.15e-31	Length:	3297
Score:	41.00	Matches:	41
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	30.83%	Indels:	0
DB:	25	Gaps:	0

US-09-638-693-36 (1-133) x ABX14411 (1-3297)

QY	58	GlyGlyLysProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMet	77
Db	2242	GGGGCAAGCCGGCAATCGTCCACACAAGAGGTGTGTATCACAAATACGATCAGATG	2301
QY	78	GluGluCysSerGlnAlaAlaProTyrTleGluGlnAlaGlnValIleAlaHisGlnPhe	97
Db	2302	GAAGAGTGTCTACAAAGCTGCCCATATATATCGAACAAGCTCAGGTAATAGCTCACCAGTTC	2361
QY	98	Lys 98	
Db	2362	ARG 2364	

Search completed: August 29, 2003, 13:01:16
Job time : 243 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run On: August 29, 2003, 12:53:25 ; Search time 58 Seconds

(without alignments)
1012.137 Million cell updates/sec

Title: US-09-638-693-36

Perfect score: 133

Sequence: 1 QNEICLHPITKYMCMSCA.....VIEPIVTTNMOKLEAFWHKH 133

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1135299

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q/cgn2_1/usptool/US09638693/runat_25082003_173341_22313/app_query.fasta_1.327
-DB=Issued_Patents_NA -QMT=fastap -SUFFIX=oli.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09638693 -CGCN_1_1_56 -runat_25082003_173341_22313 -NCPU=3
-NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents,NA:*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	36.1	367	1	US-08-244-116B-16
2	19	14.3	741	1	US-08-324-977-43
3	19	14.3	741	2	US-08-384-616-43
4	19	14.3	741	2	US-08-904-686A-43
5	19	14.3	741	3	US-09-315-850-43
6	19	14.3	1941	3	US-09-198-723A-116
7	19	14.3	1941	3	US-09-198-723A-117
8	19	14.3	1941	3	US-09-198-723A-118
9	19	14.3	1941	3	US-09-198-723A-119
10	19	14.3	1941	3	US-09-198-723A-120
11	19	14.3	1998	3	US-09-198-723A-102
12	19	14.3	1998	3	US-09-198-723A-103

13	19	14.3	1998	3	US-09-198-723A-104	Sequence 104, App
14	19	14.3	1998	3	US-09-198-723A-105	Sequence 105, App
15	19	14.3	1998	3	US-09-198-723A-106	Sequence 106, App
16	19	14.3	1998	3	US-09-198-723A-107	Sequence 107, App
17	19	14.3	1998	3	US-09-198-723A-108	Sequence 108, App
18	19	14.3	1998	3	US-09-198-723A-109	Sequence 109, App
19	19	14.3	2016	3	US-09-198-723A-110	Sequence 110, App
20	19	14.3	2016	3	US-09-198-723A-111	Sequence 111, App
21	19	14.3	6039	1	US-08-324-977-11	Sequence 11, Appl
22	19	14.3	6039	2	US-08-384-616-11	Sequence 11, Appl
23	19	14.3	6039	2	US-08-904-686A-11	Sequence 11, Appl
24	19	14.3	6039	3	US-09-315-850-11	Sequence 11, Appl
25	19	14.3	7863	1	US-08-324-977-35	Sequence 35, Appl
26	19	14.3	7863	2	US-08-384-616-35	Sequence 35, Appl
27	19	14.3	7863	2	US-08-904-686A-35	Sequence 35, Appl
28	19	14.3	7863	3	US-09-315-850-35	Sequence 35, Appl
29	19	14.3	7917	1	US-08-324-977-31	Sequence 31, Appl
30	19	14.3	7917	2	US-08-384-616-31	Sequence 31, Appl
31	19	14.3	7917	2	US-08-904-686A-31	Sequence 31, Appl
32	19	14.3	7917	3	US-09-315-850-31	Sequence 31, Appl
33	19	14.3	9030	1	US-08-324-977-13	Sequence 13, Appl
34	19	14.3	9030	2	US-08-384-616-13	Sequence 13, Appl
35	19	14.3	9030	2	US-08-904-686A-13	Sequence 13, Appl
36	19	14.3	9030	3	US-09-315-850-13	Sequence 13, Appl
37	19	14.3	9416	1	US-08-324-977-1	Sequence 1, Appl1
38	19	14.3	9416	2	US-08-384-616-1	Sequence 1, Appl1
39	19	14.3	9416	2	US-08-904-686A-1	Sequence 1, Appl1
40	19	14.3	9416	3	US-09-315-850-1	Sequence 1, Appl1
41	19	14.3	9416	4	US-08-823-895A-27	Sequence 27, Appl
42	19	14.3	9595	3	US-09-014-416-4	Sequence 4, Appl1
43	19	14.3	13910	3	US-09-263-933-1	Sequence 1, Appl1
44	19	14.3	13910	3	US-09-263-933-8	Sequence 8, Appl1
45	19	14.3	13910	3	US-09-263-933-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1

US-08-244-116B-16
; Sequence 16, Application US/08244116B
; Patent No. 5763159
; GENERAL INFORMATION:
; APPLICANT: Simmonds, Peter
; APPLICANT: Chan, Shiu-Wan
; APPLICANT: Yap, Peng L.
; TITLE OF INVENTION: Hepatitis-C Virus Testing
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell, Seltzer, Park & Gibson, P.A.
; STREET: 1211 East Morehead Street
; CITY: Charlotte
; STATE: No. 5763159th Carolina
; COUNTRY: United States
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0. Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/244,116B
; FILING DATE: 15-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB92/02143
; FILING DATE: 20-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 1749-125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 704-377-1561
; TELEFAX: 704-334-2014

```
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 367 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: CDNA
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   ORIGINAL SOURCE:
;     ORGANISM: Hepatitis-C virus
;
US-08-244-116B-16

Alignment Scores:
Pred. No.: 1.83e-40
Score: 48.00
Length: 367
Percent Similarity: 97.65%
Matches: 83
Best Local Similarity: 97.65%
Mismatches: 0
Query Match: 36.09%
Indels: 2
DB: 1
Gaps: 0

US-09-638-693-36 (1-133) x US-08-244-116B-16 (1-367)

Qy 15 MetalaCysMetSerAlaAspLeuGluValThrThrSerThrTrpValLeuLeuGlyGly 34
Db 26 AUGGCAUGYAGUGCAGCUGAUCUGGAGAAAGUAACCAACCAACGCGGUGGUGGAGGR 85
Qy 35 ValLeuAlaLeuAlaAlaTyrcysLeuSerValGlyCysValValIleValGlyHis 54
Db 86 GUCCUCGKCCUGGCGGCUACUGCUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGG 145
Qy 55 IleGluLeuGlyGlyLysProAlaIle-ValProAspLysGluValLeuTyrcysGln 74
Db 146 AUGGAGCUGGCGGCAAGCCVGCAMU-CGUUCCAGACAARGARGUGUUAUCAAUA 204
Qy 74 rAspGluMetGluCysSerGlnAlaAlaProTyrcysIleGluGlnAlaGlnValIleAl 94
Db 205 CGAUGAGAGGAGGAGGUGGUGGAGGUGGUGGAGGUGGUGGAGGUGGUGGAGGUGG 264
Qy 94 aHisGlnPhelys 98
Db 265 CCACCAAGUUAAG 277

RESULT 2
US-08-324-977-43
; Sequence 43, Application US/08324977
; Patent No. 5747339
; GENERAL INFORMATION:
;   APPLICANT: OKAYAMA, Hiroto
;   APPLICANT: FUKU, Isao
;   APPLICANT: MORI, Chisato
;   APPLICANT: TAKAMIZAWA, Akahisa
;   APPLICANT: YOSHIDA, Iwao
;   TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
;   TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
;   NUMBER OF SEQUENCES: 50
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Armstrong, Westernman, Hattori, McLeland &
;     STREET: 1725 K St. N.W. Suite 1000
;     CITY: Washington
;     STATE: D.C.
;     COUNTRY: U.S.A.
;     ZIP: 20006
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette, 3.5 in, 1.44MB
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
;     SOFTWARE: ASCII
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/324,977
;     FILING DATE: 18-OCT-1994
;   PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: JP 2-167466
; FILING DATE: 25-JUN-1990
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: JP 2-230921
;   FILING DATE: 31-AUG-1990
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: JP 2-305605
;   FILING DATE: 09-NOV-1990
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/099,706
;   FILING DATE: 30-JUL-1993
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/769,996
;   FILING DATE: 02-OCT-1991
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/635,451
;   FILING DATE: 28-DEC-1990
; ATTORNEY/AGENT INFORMATION:
;   NAME: Stevens-Smith, Theresa M.
;   REGISTRATION NUMBER: 36,281
;   REFERENCE/DOCKET NUMBER: 900703D
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (202) 659-2930
;   TELEFAX: (202) 887-0357
;   TELEX: 440142
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 741 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: CDNA from genomic RNA
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 1..741
;     OTHER INFORMATION: /note: "sequence = 5178 - 5918 of
;     OTHER INFORMATION: SEQ ID NO: 1"
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 1..741
; US-08-324-977-43

Alignment Scores:
Pred. No.: 1.26e-10
Score: 19.00
Length: 741
Percent Similarity: 100.00%
Matches: 19
Best Local Similarity: 100.00%
Mismatches: 0
Query Match: 14.29%
Indels: 0
DB: 1
Gaps: 0

US-09-638-693-36 (1-133) x US-08-324-977-43 (1-741)

Qy 6 LeuThrHisProIleThrLysTyrcysMetSerAlaAspLeuGluVal 24
Db 64 CTCACCCACCCCAATACCAATATCATCATGCGTGCCTGACCTGGAGGTC 120

RESULT 3
US-08-384-616-43
; Sequence 43, Application US/08384616
; Patent No. 5847101
; GENERAL INFORMATION:
;   APPLICANT: OKAYAMA, Hiroto
;   APPLICANT: FUKU, Isao
;   APPLICANT: MORI, Chisato
;   APPLICANT: TAKAMIZAWA, Akahisa
;   APPLICANT: YOSHIDA, Iwao
;   TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
;   TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
;   NUMBER OF SEQUENCES: 50
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Armstrong, Westernman, Hattori, McLeland &
;     STREET: 1725 K St. N.W. Suite 1000
```

APPLICANT: MORI, Chisato
APPLICANT: TAKAMIZAWA, Akahisa
APPLICANT: YOSHIDA, Iwao
TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Armstrong, Westerman, Hattori, Mcleland &
ADDRESS: Naughton
STREET: 1725 K St. N.W. Suite 1000
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 in, 1.44Mb
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/384,616
FILING DATE: 02-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-167466
FILING DATE: 25-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-230921
FILING DATE: 31-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 2-305605
FILING DATE: 09-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/635,451
FILING DATE: 28-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Stevens-Smith, Theresa M.
REGISTRATION NUMBER: 36,281
REFERENCE/DOCKET NUMBER: 900703B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 659-2930
TELEFAX: (202) 887-0357
TELEX: 440142
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 741 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA from genomic RNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..741
OTHER INFORMATION: /note: "sequence - 5178 - 5918 of
OTHER INFORMATION: SEQ ID NO: 1"
FEATURE:
NAME/KEY: CDS
LOCATION: 1..741
US-08-384-616-43

Alignment Scores:
Pred. No.: 1.26e-10 Length: 741
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.29% Indels: 0
DB: 2 Gaps: 0

US-09-638-693-36 (1-133) x US-08-384-616-43 (1-741)

QY 6 LeuThrHisProIleThrIleMetAlaCysMetSerAlaAspLeuGluVal 24
Db 64 CTCACCCACCCATAACCAATACATCATGCGTCGACCTGGAGGTC 120

RESULT 4

US-08-904-686A-43
Sequence 43, Application US/08904686A
Patent No. 5998130
GENERAL INFORMATION:
APPLICANT: OKAYAMA, Hiroto
APPLICANT: FUKU, Isao

Alignment Scores:
Pred. No.: 1.26e-10 Length: 741
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.29% Indels: 0
US-08-904-686A-43

```
DB:                2                Gaps:                0
US-09-638-693-36 (1-133) x US-08-904-686A-43 (1-741)
Qy      6 LeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAlaAspLeuGluVal 24
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64 CTCACCCACCCATAACCAATACATCATGCGCATGTCGCGCTGACCTGGAGGTC 120

RESULT 5
US-09-315-850-43
; Sequence 43, Application US/09315850
; Patent No. 6217872
; GENERAL INFORMATION:
; APPLICANT: OKAYAMA, Hiroto
; APPLICANT: FUKU, Isao
; APPLICANT: MORI, Chisato
; APPLICANT: TAKAMIZAWA, Akahisa
; APPLICANT: YOSHIDA, Iwao
; TITLE OF INVENTION: NON-A, NON-B HEPATITIS VIRUS GENOMIC
; TITLE OF INVENTION: CDNA AND ANTIGEN POLYPEPTIDE
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Armstrong, Westerman, Hattori, McLeLend &
; ADDRESSEE: Naughton
; STREET: 1725 K St. N.W. Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 in, 1.44Mb
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS, Version 5.0
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/315,850
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/904,686
; FILING DATE: 01-AUG-1997
; APPLICATION NUMBER: US 08/324,977
; FILING DATE: 18-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-167466
; FILING DATE: 25-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-230921
; FILING DATE: 31-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 2-305605
; FILING DATE: 09-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/099,706
; FILING DATE: 30-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/769,996
; FILING DATE: 02-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/635,451
; FILING DATE: 28-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: McLeLend, Le-Nhung
; REGISTRATION NUMBER: 31,541
; REFERENCE/DOCKET NUMBER: 900703G
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 659-2930
; TELEFAX: (202) 887-0357
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 741 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: cdna from genomic RNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..741
; OTHER INFORMATION: /note: "sequence = 5178 - 5918 of
; OTHER INFORMATION: SEQ ID NO: 1"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..741
US-09-315-850-43
Alignment Scores:
Pred. NO.: 1.26e-10 Length: 741
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.29% Indels: 0
DB: 3 Gaps: 0
US-09-638-693-36 (1-133) x US-09-315-850-43 (1-741)
Qy      6 LeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAlaAspLeuGluVal 24
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64 CTCACCCACCCATAACCAATACATCATGCGCATGTCGCGCTGACCTGGAGGTC 120

RESULT 6
US-09-198-723A-116
; Sequence 116, Application US/09198723A
; Patent No. 6211338
; GENERAL INFORMATION:
; APPLICANT: Malcolm, Bruce
; APPLICANT: Taremi, Shahriar S.
; APPLICANT: Weber, Patricia
; APPLICANT: Yao, Nahua
; TITLE OF INVENTION: Covalent Complexes of Hepatitis C Virus
; TITLE OF INVENTION: NS3 Protease and NS4A Cofactor Peptide
; NUMBER OF SEQUENCES: 123
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-Plough Corp.
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07030
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Power Macintosh
; OPERATING SYSTEM: 8.0.1
; SOFTWARE: Microsoft Word 6.0.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/198,723A
; FILING DATE: 24 NOV 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McLaughlin, Jaye P.
; REGISTRATION NUMBER: 41,211
; REFERENCE/DOCKET NUMBER: J080800
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)298-5056
; TELEFAX: (908)298-5388
; INFORMATION FOR SEQ ID NO: 116:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1941 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1941
US-09-198-723A-116
Alignment Scores:
```

Pred. No.: 3.12e-10 Length: 1941
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.29% Indels: 0
DB: 3 Gaps: 0

US-09-638-693-36 (1-133) x US-09-198-723A-116 (1-1941)

QY 6 LeuThrHisProfilThrLysTyrIleMetAlaCysMetSerAlaAspLeuGluVal 24
|||||
Db 1876 CTCACCCACCCATACCAATAACATCATGGCATGTCGCCCGACCTGGAGGTC 1932

RESULT 7

US-09-198-723A-117
; Sequence 117, Application US/09198723A
; Patent No. 6211338
; GENERAL INFORMATION:
; APPLICANT: Malcolm, Bruce
; APPLICANT: Taremi, Shahrar S.
; APPLICANT: Weber, Patricia
; APPLICANT: Yao, Nanhua
; TITLE OF INVENTION: Covalent Complexes of Hepatitis C Virus
; TITLE OF INVENTION: NS3 Protease and NS4A Cofactor Peptide
; NUMBER OF SEQUENCES: 123
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-Plough Corp.
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07030

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Power Macintosh
OPERATING SYSTEM: 8.0.1
SOFTWARE: Microsoft Word 6.0.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/198,723A
FILING DATE: 24 NOV 1998

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: McLaughlin, Jaye P.
REGISTRATION NUMBER: 41,211
REFERENCE/DOCKET NUMBER: JB0800
TELEPHONE: (908)298-5056
TELEFAX: (908)298-5388

INFORMATION FOR SEQ ID NO: 117:
SEQUENCE CHARACTERISTICS:
LENGTH: 1941 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1941

US-09-198-723A-117

Alignment Scores:
Pred. No.: 3.12e-10 Length: 1941
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.29% Indels: 0
DB: 3 Gaps: 0

US-09-638-693-36 (1-133) x US-09-198-723A-117 (1-1941)

QY 6 LeuThrHisProfilThrLysTyrIleMetAlaCysMetSerAlaAspLeuGluVal 24
|||||
Db 1876 CTCACCCACCCATACCAATAACATCATGGCATGTCGCCCGACCTGGAGGTC 1932

RESULT 8

US-09-198-723A-118
; Sequence 118, Application US/09198723A
; Patent No. 6211338
; GENERAL INFORMATION:
; APPLICANT: Malcolm, Bruce
; APPLICANT: Taremi, Shahrar S.
; APPLICANT: Weber, Patricia
; APPLICANT: Yao, Nanhua

; TITLE OF INVENTION: Covalent Complexes of Hepatitis C Virus
; TITLE OF INVENTION: NS3 Protease and NS4A Cofactor Peptide
; NUMBER OF SEQUENCES: 123
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-Plough Corp.
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07030

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Power Macintosh
OPERATING SYSTEM: 8.0.1
SOFTWARE: Microsoft Word 6.0.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/198,723A
FILING DATE: 24 NOV 1998

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: McLaughlin, Jaye P.
REGISTRATION NUMBER: 41,211
REFERENCE/DOCKET NUMBER: JB0800
TELEPHONE: (908)298-5056
TELEFAX: (908)298-5388

INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 1941 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1941

US-09-198-723A-118

Alignment Scores:
Pred. No.: 3.12e-10 Length: 1941
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.29% Indels: 0
DB: 3 Gaps: 0

US-09-638-693-36 (1-133) x US-09-198-723A-118 (1-1941)

QY 6 LeuThrHisProfilThrLysTyrIleMetAlaCysMetSerAlaAspLeuGluVal 24
|||||
Db 1876 CTCACCCACCCATACCAATAACATCATGGCATGTCGCCCGACCTGGAGGTC 1932

RESULT 9

US-09-198-723A-119
; Sequence 119, Application US/09198723A
; Patent No. 6211338
; GENERAL INFORMATION:
; APPLICANT: Malcolm, Bruce
; APPLICANT: Taremi, Shahrar S.
; APPLICANT: Weber, Patricia
; APPLICANT: Yao, Nanhua

; TITLE OF INVENTION: Covalent Complexes of Hepatitis C Virus
; TITLE OF INVENTION: NS3 Protease and NS4A Cofactor Peptide

```

; NUMBER OF SEQUENCES: 123
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-Plough Corp.
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07030
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Power Macintosh
; OPERATING SYSTEM: 8.0.1
; SOFTWARE: Microsoft Word 6.0.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/198,723A
; FILING DATE: 24 NOV 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McLaughlin, Jaye P.
; REGISTRATION NUMBER: 41,211
; REFERENCE/DOCKET NUMBER: JB0800
; TELEPHONE: (908)298-5056
; TELEFAX: (908)298-5388
; INFORMATION FOR SEQ ID NO: 119:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1941 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1941
; US-09-198-723A-119
;
; Alignment Scores:
; Pred. No.: 3.12e-10 Length: 1941
; Score: 19.00 Matches: 19
; Percent Similarity: 100.00% Conservative: 0
; Best Local Similarity: 100.00% Mismatches: 0
; Query Match: 14.29% Indels: 0
; DB: 3 Gaps: 0
;
; US-09-638-693-36 (1-133) x US-09-198-723A-119 (1-1941)
;
; Qy 6 LeuThrHisProLeuThrLysTyrIleMetalaCysMetSerAlaAspLeuGluVal 24
; Db 1876 CTCACCCACCCCAATACAAATACATCATGCGCGGACCTGGAGGTC 1932
;
; RESULT 10
; US-09-198-723A-120
; Sequence 120, Application US/09198723A
; Patent No. 6211338
; GENERAL INFORMATION:
; APPLICANT: Malcolin, Bruce
; APPLICANT: Taremi, Shahriar S.
; APPLICANT: Weber, Patricia
; APPLICANT: Yao, Nannhua
; TITLE OF INVENTION: Covalent Complexes of Hepatitis C Virus
; NUMBER OF SEQUENCES: 123
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-Plough Corp.
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07030
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Power Macintosh
; OPERATING SYSTEM: 8.0.1
;
; US-09-198-723A-102
; Sequence 102, Application US/09198723A
; Patent No. 6211338
; GENERAL INFORMATION:
; APPLICANT: Malcolin, Bruce
; APPLICANT: Taremi, Shahriar S.
; APPLICANT: Weber, Patricia
; APPLICANT: Yao, Nannhua
; TITLE OF INVENTION: Covalent Complexes of Hepatitis C Virus
; NUMBER OF SEQUENCES: 123
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-Plough Corp.
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07030
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Power Macintosh
; OPERATING SYSTEM: 8.0.1
; SOFTWARE: Microsoft Word 6.0.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/198,723A
; FILING DATE: 24 NOV 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McLaughlin, Jaye P.
; REGISTRATION NUMBER: 41,211
; REFERENCE/DOCKET NUMBER: JB0800
; TELEPHONE: (908)298-5056
; TELEFAX: (908)298-5388

```

```

; SOFTWARE: Microsoft Word 6.0.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/198,723A
; FILING DATE: 24 NOV 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McLaughlin, Jaye P.
; REGISTRATION NUMBER: 41,211
; REFERENCE/DOCKET NUMBER: JB0800
; TELEPHONE: (908)298-5056
; TELEFAX: (908)298-5388
; INFORMATION FOR SEQ ID NO: 120:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1941 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1941
; US-09-198-723A-120
;
; Alignment Scores:
; Pred. No.: 3.12e-10 Length: 1941
; Score: 19.00 Matches: 19
; Percent Similarity: 100.00% Conservative: 0
; Best Local Similarity: 100.00% Mismatches: 0
; Query Match: 14.29% Indels: 0
; DB: 3 Gaps: 0
;
; US-09-638-693-36 (1-133) x US-09-198-723A-120 (1-1941)
;
; Qy 6 LeuThrHisProLeuThrLysTyrIleMetalaCysMetSerAlaAspLeuGluVal 24
; Db 1876 CTCACCCACCCCAATACAAATACATCATGCGCGGACCTGGAGGTC 1932
;
; RESULT 11
; US-09-198-723A-102
; Sequence 102, Application US/09198723A
; Patent No. 6211338
; GENERAL INFORMATION:
; APPLICANT: Malcolin, Bruce
; APPLICANT: Taremi, Shahriar S.
; APPLICANT: Weber, Patricia
; APPLICANT: Yao, Nannhua
; TITLE OF INVENTION: Covalent Complexes of Hepatitis C Virus
; NUMBER OF SEQUENCES: 123
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-Plough Corp.
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07030
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Power Macintosh
; OPERATING SYSTEM: 8.0.1
; SOFTWARE: Microsoft Word 6.0.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/198,723A
; FILING DATE: 24 NOV 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McLaughlin, Jaye P.
; REGISTRATION NUMBER: 41,211
; REFERENCE/DOCKET NUMBER: JB0800
; TELEPHONE: (908)298-5056
; TELEFAX: (908)298-5388

```


; INFORMATION FOR SEQ ID NO: 102:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1998 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1998

US-09-198-723A-102

Alignment Scores:
Pred. No.: 3.2e-10 Length: 1998
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.29% Indels: 0
DB: 3 Gaps: 0

US-09-638-693-36 (1-133) x US-09-198-723A-102 (1-1998)

QY 6 LeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAlaAspLeuGluVal 24
|||||
Db 1936 CTCACCCACCCATAACCAATACATCATGGCATGCGGCTGACCTGGAGGTC 1992

RESULT 12

US-09-198-723A-103
; Sequence 103, Application US/09198723A
; Patent No. 6211338

; GENERAL INFORMATION:

; APPLICANT: Malcolm, Bruce
; APPLICANT: Taremi, Shahriar S.
; APPLICANT: Weber, Patricia
; APPLICANT: Yao, Nanhua

; TITLE OF INVENTION: Covalent Complexes of Hepatitis C Virus
; NUMBER OF SEQUENCES: 123
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-Plough Corp.
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07030

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: Power Macintosh
; OPERATING SYSTEM: 8.0.1
; SOFTWARE: Microsoft Word 6.0.1

; CURRENT APPLICATION DATA: 123

; APPLICATION NUMBER: US/09/198,723A
; FILING DATE: 24 NOV 1998

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: McLaughlin, Jaye P.
; REGISTRATION NUMBER: 41,211
; REFERENCE/DOCKET NUMBER: JB0800

; TELEPHONE: (908)298-5056

; TELEFAX: (908)298-5388

; INFORMATION FOR SEQ ID NO: 103:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1998 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:

; NAME/KEY: CDS
; LOCATION: 1..1998

US-09-198-723A-103

Alignment Scores:

Pred. No.: 3.2e-10 Length: 1998
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.29% Indels: 0
DB: 3 Gaps: 0

US-09-638-693-36 (1-133) x US-09-198-723A-103 (1-1998)

QY 6 LeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAlaAspLeuGluVal 24
|||||
Db 1936 CTCACCCACCCATAACCAATACATCATGGCATGCGGCTGACCTGGAGGTC 1992

RESULT 13

US-09-198-723A-104
; Sequence 104, Application US/09198723A
; Patent No. 6211338

; GENERAL INFORMATION:

; APPLICANT: Malcolm, Bruce
; APPLICANT: Taremi, Shahriar S.
; APPLICANT: Weber, Patricia
; APPLICANT: Yao, Nanhua

; TITLE OF INVENTION: Covalent Complexes of Hepatitis C Virus
; NUMBER OF SEQUENCES: 123
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schering-Plough Corp.
; STREET: 2000 Galloping Hill Road
; CITY: Kenilworth
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07030

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: Power Macintosh
; OPERATING SYSTEM: 8.0.1
; SOFTWARE: Microsoft Word 6.0.1

; CURRENT APPLICATION DATA: US/09/198,723A
; APPLICATION NUMBER: 24 NOV 1998

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: McLaughlin, Jaye P.
; REGISTRATION NUMBER: 41,211
; REFERENCE/DOCKET NUMBER: JB0800

; TELEPHONE: (908)298-5056

; TELEFAX: (908)298-5388

; INFORMATION FOR SEQ ID NO: 104:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1998 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..1998

US-09-198-723A-104

Alignment Scores:

Pred. No.: 3.2e-10 Length: 1998
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.29% Indels: 0
DB: 3 Gaps: 0

US-09-638-693-36 (1-133) x US-09-198-723A-104 (1-1998)

QY 6 LeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAlaAspLeuGluVal 24
|||||

Db 1936 CTCACCCACCCATAACCAATACATCATGCGTGCATGCGGTGACCTGGAGGTC 1992

RESULT 14

US-09-198-723A-105

; Sequence 105, Application US/09198723A

; Patent No. 6211338

; GENERAL INFORMATION:

; APPLICANT: Malcolm, Bruce

; APPLICANT: Taremi, Shahrar S.

; APPLICANT: Weber, Patricia

; APPLICANT: Yao, Nanhua

; TITLE OF INVENTION: Covalent Complexes of Hepatitis C Virus

; NUMBER OF SEQUENCES: 123

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Schering-Plough Corp.

; STREET: 2000 Galloping Hill Road

; CITY: Kenilworth

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07030

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: Power Macintosh

; OPERATING SYSTEM: 8.0.1

; SOFTWARE: Microsoft Word 6.0.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/198,723A

; FILING DATE: 24 NOV 1998

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: McLaughlin, Jaye P.

; REGISTRATION NUMBER: 41,211

; REFERENCE/DOCKET NUMBER: JB0800

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (908)298-5056

; TELEFAX: (908)298-5388

; INFORMATION FOR SEQ ID NO: 105:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1998 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..1998

US-09-198-723A-105

Alignment Scores:

Pred. No.:	3.2e-10	Length:	1998
Score:	19.00	Matches:	19
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	14.29%	Indels:	0
DB:	3	Gaps:	0

US-09-638-693-36 (1-133) x US-09-198-723A-105 (1-1998)

Qy 6 LeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAlaAspLeuGluVal 24

Db 1936 CTCACCCACCCATAACCAATACATCATGCGTGCATGCGGTGACCTGGAGGTC 1992

RESULT 15

US-09-198-723A-106

; Sequence 106, Application US/09198723A

; Patent No. 6211338

; GENERAL INFORMATION:

; APPLICANT: Malcolm, Bruce

; APPLICANT: Taremi, Shahrar S.

; APPLICANT: Weber, Patricia

; APPLICANT: Yao, Nanhua

; TITLE OF INVENTION: Covalent Complexes of Hepatitis C Virus

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 29, 2003, 12:57:10 ; Search time 215 Seconds

(without alignments)
1419.234 Million cell updates/sec

Title: US-09-638-693-36

Perfect score: 133

Sequence: 1 ONEICLTHPTIKYIMACMSA.....VIEPIVTTNWKLEAFNKH 133

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1533700 seqs, 1147125425 residues

Word size: 1

Total number of hits satisfying chosen parameters: 3063161

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q/cgn2_1/USPTO_spool/US09638693/runat_25082003_173341_22326/app_query.fasta_1.327
-DB=PublishedApplications_NA -QFMT=fastap -SUFFIX=oli.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USRP=US09638693 @CGN_1_1_221 @runat_25082003_173341_22326
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Published Applications_NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	133	100.0	401	11	US-09-899-046-35
					Sequence 35, Appl

2	133	100.0	401	11	US-09-878-281-35	Sequence 35, Appl
3	98	73.7	401	11	US-09-899-046-37	Sequence 37, Appl
4	98	73.7	401	11	US-09-899-046-39	Sequence 39, Appl
5	98	73.7	401	11	US-09-878-281-37	Sequence 37, Appl
6	98	73.7	401	11	US-09-878-281-39	Sequence 39, Appl
7	98	73.7	401	11	US-09-899-046-222	Sequence 222, Appl
8	98	73.7	629	11	US-09-878-281-222	Sequence 222, Appl
9	52	39.1	401	11	US-09-899-046-31	Sequence 31, Appl
10	52	39.1	401	11	US-09-899-046-33	Sequence 33, Appl
11	52	39.1	401	11	US-09-878-281-31	Sequence 31, Appl
12	52	39.1	401	11	US-09-878-281-33	Sequence 33, Appl
13	41	30.8	2499	10	US-09-881-239-4	Sequence 4, Appl
14	41	30.8	3297	10	US-09-881-654-3	Sequence 3, Appl
15	19	14.3	287	11	US-09-899-046-29	Sequence 29, Appl
16	19	14.3	287	11	US-09-878-281-29	Sequence 29, Appl
17	19	14.3	1443	11	US-09-899-046-269	Sequence 269, Appl
18	19	14.3	1443	11	US-09-878-281-269	Sequence 269, Appl
19	19	14.3	1485	11	US-09-899-046-197	Sequence 197, Appl
20	19	14.3	1485	11	US-09-899-046-199	Sequence 199, Appl
21	19	14.3	1485	11	US-09-878-281-197	Sequence 197, Appl
22	19	14.3	1485	11	US-09-878-281-199	Sequence 199, Appl
23	19	14.3	9275	14	US-10-259-275-39	Sequence 39, Appl
24	19	14.3	13910	11	US-09-919-901-1	Sequence 1, Appl
25	19	14.3	13910	11	US-09-919-901-8	Sequence 8, Appl
26	19	14.3	13910	11	US-09-919-901-15	Sequence 15, Appl
27	18	13.5	6189	14	US-10-259-275-41	Sequence 41, Appl
28	18	13.5	7992	13	US-10-005-469-1	Sequence 1, Appl
29	18	13.5	7992	13	US-10-005-469-2	Sequence 2, Appl
30	18	13.5	7992	13	US-10-005-469-4	Sequence 4, Appl
31	18	13.5	7992	13	US-10-005-469-5	Sequence 5, Appl
32	18	13.5	7992	13	US-10-005-469-6	Sequence 6, Appl
33	18	13.5	7995	13	US-10-005-469-3	Sequence 3, Appl
34	18	13.5	8638	12	US-10-309-561-6	Sequence 6, Appl
35	18	13.5	8638	12	US-10-309-561-7	Sequence 7, Appl
36	18	13.5	8638	12	US-10-309-561-24	Sequence 24, Appl
37	18	13.5	8638	12	US-10-309-561-25	Sequence 25, Appl
38	18	13.5	8638	13	US-10-029-907-6	Sequence 6, Appl
39	18	13.5	8638	13	US-10-029-907-7	Sequence 7, Appl
40	18	13.5	8638	13	US-10-029-907-24	Sequence 24, Appl
41	18	13.5	8638	13	US-10-029-907-25	Sequence 25, Appl
42	18	13.5	8639	12	US-10-309-561-1	Sequence 1, Appl
43	18	13.5	8639	13	US-10-029-907-1	Sequence 1, Appl
44	18	13.5	8642	12	US-10-309-561-2	Sequence 2, Appl
45	18	13.5	8642	13	US-10-029-907-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-09-899-046-35
; Sequence 35, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 401 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

١٤٤٠

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: BR36-20-166
FEATURE:
NAME/KEY: CDS
LOCATION: 3..401
US-09-899-046-37

Alignment Scores:
Pred. No.: 1,04e-96 Length: 401
Score: 98.00 Matches: 132
Percent Similarity: 98.51% Conservatives: 0
Best Local Similarity: 98.51% Mismatches: 1
Query Match: 73.68% Indels: 2
DB: 11 Gaps: 0

US-09-638-693-36 (1-133) x US-09-899-046-37 (1-401)

QY 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
|||||
Db 3 CAAATGAATCTGCTTGACACACCCCATCACAAATACATCATGGCATGTCAGCT 62
QY 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLeuAla 40
|||||
Db 63 GATCTGGAAGTAACCAACACAGCACCTGGGTTTCTTGGAGGGTCTCGCGCCCTAGCG 122
QY 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyGlyLys 60
|||||
Db 123 GCCTACTGCTTGCAGTCGGTTGTGTGATGTGGTCATATCGAGCTGGGGGCAAG 182
QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGlyCys 80
|||||
Db 183 CCGGCAATCGTTCACACAAAGAGGTGTGTATCAACAATACGATGAGATGGAAGAGTGC 242
QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPhelLysGly-Ly 100
|||||
Db 243 TCACAGCTGCCCATATATCGAAGCTCAGGTAGTACGATGAGTCAACAGTTCAA-GGAAAA 301
QY 100 sValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValTh 120
Db 302 AGTCCTTGGATTGCTGCAGCGAGCCACCAACAAGCTGTCAITGAGCCCATAGTAAC 361
QY 120 rThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
|||||
Db 362 TACCAACTGGCAAAAGCTTGAGGCCTTTTGGCACAAGCAT 401

RESULT 4

US-09-899-046-37
Sequence 39, Application US/09899046
Publication No. US2003008274A1
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: New sequences of hepatitis C virus

TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
NUMBER OF SEQUENCES: 270
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,046
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/362,455
FILING DATE:
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:

LENGTH: 401 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: BR36-20-165
FEATURE:
NAME/KEY: CDS
LOCATION: 3..401
US-09-899-046-39

Alignment Scores:
Pred. No.: 1,04e-96 Length: 401
Score: 98.00 Matches: 132
Percent Similarity: 98.51% Conservatives: 0
Best Local Similarity: 98.51% Mismatches: 1
Query Match: 73.68% Indels: 2
DB: 11 Gaps: 0

US-09-638-693-36 (1-133) x US-09-899-046-39 (1-401)

QY 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
|||||
Db 3 CAAATGAATCTGCTTGACACACCCCATCACAAATACATCATGGCATGTCAGCT 62
QY 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLeuAla 40
|||||
Db 63 GATCTGGAAGTAACCAACACAGCACCTGGGTTTCTTGGAGGGTCTCGCGCCCTAGCG 122
QY 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyGlyLys 60
|||||
Db 123 GCCTACTGCTTGCAGTCGGTTGTGTGATGTGGTCATATCGAGCTGGGGGCAAG 182
QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGlyCys 80
|||||
Db 183 CCGGCAATCGTTCACACAAAGAGGTGTGTATCAACAATACGATGAGATGGAAGAGTGC 242
QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPhelLysGly-Ly 100
|||||
Db 243 TCACAGCTGCCCATATATCGAAGCTCAGGTAGTACGATGAGTCAACAGTTCAA-GGAAAA 301
QY 100 sValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValTh 120
Db 302 AGTCCTTGGATTGCTGCAGCGAGCCACCAACAAGCTGTCAITGAGCCCATAGTAAC 361
QY 120 rThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
|||||
Db 362 TACCAACTGGCAAAAGCTTGAGGCCTTTTGGCACAAGCAT 401

RESULT 5

US-09-878-281-37
Sequence 37, Application US/09878281
Publication No. US20030032005A1
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: New sequences of hepatitis C virus

TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
NUMBER OF SEQUENCES: 270
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/878,281
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/362,455
FILING DATE:
INFORMATION FOR SEQ ID NO: 37:

SEQUENCE CHARACTERISTICS:
LENGTH: 401 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: BR36-20-166
FEATURE:
NAME/KEY: CDS
LOCATION: 3..401
US-09-878-281-37

Alignment Scores:
Pred. No.: 1.04e-96 Length: 401
Score: 98.00 Matches: 132
Percent Similarity: 98.51% Conservative: 0
Best Local Similarity: 98.51% Mismatches: 1
Query Match: 73.68% Indels: 2
DB: 11 Gaps: 0

US-09-638-693-36 (1-133) x US-09-878-281-37 (1-401)

Qy 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTrpIleMetAlaCysMetSerAla 20
|||||
Db 3 CAAATGAATCTGCTTGACACACCCCATCACAATAATACATCATGCGATGTCAGCT 62
Qy 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyValLeuAlaLeuAla 40
|||||
Db 63 GATCGAAGTAACACACACACCTGGGTTTGTGGAGGGGTCCTCGCGGCCCTAGCG 122
Qy 41 AlaTrpCysLeuSerValGlyCysValIleValGlyHisIleGluLeuGlyGly 60
|||||
Db 123 GCCTACTGCTGTGTCAGTCGGTTGTGTGATGTTGGTGTATCATCGAGCTGGGGGCAAG 182
Qy 61 ProAlaIleValProAspLysGluValLeuTrpGlnGlnTrpAspGluMetGluGluCys 80
|||||
Db 183 CCGCAATCGTTCAGACAAAGAGGTGTTGATCAACAATACGATGAGATGGAAGAGTGC 242
Qy 81 SerGlnAlaAlaProTrpIleGluGlnAlaGlnValIleAlaHisGlnPheLysGly-Ly 100
|||||
Db 243 TCACAAGCTGCCCATATATCGAACAAAGCTCAGGTGATGCTCACCAGTTCAA-GGAAAA 301
Qy 100 sValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValTh 120
|||||
Db 302 AGTCCTGGATTGCTGCGAGGAGCCACCCACAAAGCTGTCTATTGAGCCCATAGTAAC 361
Qy 120 rThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
|||||
Db 362 TACCAACTGGCAAAAGCTTGAGGCTTTTGGCACAAAGCAT 401

RESULT 6

US-09-878-281-39

Sequence 39, Application US/09878281

Publication No. US20030032005A1

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: New sequences of hepatitis C virus

TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.

NUMBER OF SEQUENCES: 270

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION NUMBER: US/09/878,281

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/362,455

FILING DATE:

INFORMATION FOR SEQ ID NO: 39:

SEQUENCE CHARACTERISTICS:

LENGTH: 401 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

IMMEDIATE SOURCE:

CLONE: BR36-20-165

FEATURE:

NAME/KEY: CDS

LOCATION: 3..401

US-09-878-281-39

Alignment Scores:

Pred. No.: 1.04e-96 Length: 401

Score: 98.00 Matches: 132

Percent Similarity: 98.51% Conservative: 0

Best Local Similarity: 98.51% Mismatches: 1

Query Match: 73.68% Indels: 2

DB: 11 Gaps: 0

US-09-638-693-36 (1-133) x US-09-878-281-39 (1-401)

Qy 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTrpIleMetAlaCysMetSerAla 20
|||||
Db 3 CAAATGAATCTGCTTGACACACCCCATCACAATAATACATCATGCGATGTCAGCT 62
Qy 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyValLeuAlaLeuAla 40
|||||
Db 63 GATCGAAGTAACACACACACCTGGGTTTGTGGAGGGGTCCTCGCGGCCCTAGCG 122
Qy 41 AlaTrpCysLeuSerValGlyCysValIleValGlyHisIleGluLeuGlyGly 60
|||||
Db 123 GCCTACTGCTGTGTCAGTCGGTTGTGTGATGTTGGTGTATCATCGAGCTGGGGGCAAG 182
Qy 61 ProAlaIleValProAspLysGluValLeuTrpGlnGlnTrpAspGluMetGluGluCys 80
|||||
Db 183 CCGCAATCGTTCAGACAAAGAGGTGTTGATCAACAATACGATGAGATGGAAGAGTGC 242
Qy 81 SerGlnAlaAlaProTrpIleGluGlnAlaGlnValIleAlaHisGlnPheLysGly-Ly 100
|||||
Db 243 TCACAAGCTGCCCATATATCGAACAAAGCTCAGGTGATGCTCACCAGTTCAA-GGAAAA 301
Qy 100 sValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValTh 120
|||||
Db 302 AGTCCTGGATTGCTGCGAGGAGCCACCCACAAAGCTGTCTATTGAGCCCATAGTAAC 361
Qy 120 rThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
|||||
Db 362 TACCAACTGGCAAAAGCTTGAGGCTTTTGGCACAAAGCAT 401

RESULT 7

US-09-899-046-222

Sequence 222, Application US/09899046

Publication No. US20030008274A1

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: New sequences of hepatitis C virus

TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.

NUMBER OF SEQUENCES: 270

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/899,046

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/362,455

;; FILING DATE:
;; INFORMATION FOR SEQ ID NO: 222:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 629 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; ANTI-SENSE: NO
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 3..629
;; NAME/KEY: mat_peptide
;; LOCATION: 3..629
US-09-899-046-222

Alignment Scores:
Pred. No.: 1.57e-96 Length: 629
Score: 98.00 Matches: 132
Percent Similarity: 98.51% Conservative: 0
Best Local Similarity: 98.51% Mismatches: 1
Query Match: 73.68% Indels: 2
DB: 11 Gaps: 0

US-09-638-693-36 (1-133) x US-09-899-046-222 (1-629)

QY 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
|||||
Db 231 CAAATGAATCGTTGACACACCCCATCACAAAATACATCGCATGTCAGCT 290
QY 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLeuAla 40
|||||
Db 291 GATCTGGAGTAACCCAGCACCCTGGGTTTGTCTGGAGGGGTCTCGCGCCCTAGCG 350
QY 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyLys 60
|||||
Db 351 GCCTACTGCTTCAGTCGGTTGTGTGTGATGTGGTCAATCGAGCTGGGGGCAAG 410
QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluCys 80
|||||
Db 411 CGGCAATGTTCCAGACAAAGAGGTGTGTATCAACAATACGATGAGTGAAGAGTGC 470
QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGly-Ly 100
|||||
Db 471 TCACAAGCTGCCCATATATCGAAGCTCAGGTAAAGTCAAGTCAAGTCAA-GGAAAA 529
QY 100 sValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValTh 120
|||||
Db 530 AGTCCTGGATTGCTGCAGCAGGAGCCACCAACAAGCTGTCATTGAGCCCATAGTAA 589
QY 120 rThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
|||||
Db 590 TACCAACTGGCAAAAGCTTGAGGCGCTTTTGGCACAAGCAT 629

RESULT 8

US-09-878-281-222
; Sequence 222, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/878,281
; FILING DATE:

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/362,455
;; FILING DATE:
;; INFORMATION FOR SEQ ID NO: 222:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 629 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 3..629
;; NAME/KEY: mat_peptide
;; LOCATION: 3..629
US-09-878-281-222

Alignment Scores:
Pred. No.: 1.57e-96 Length: 629
Score: 98.00 Matches: 132
Percent Similarity: 98.51% Conservative: 0
Best Local Similarity: 98.51% Mismatches: 1
Query Match: 73.68% Indels: 2
DB: 11 Gaps: 0

US-09-638-693-36 (1-133) x US-09-878-281-222 (1-629)

QY 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
|||||
Db 231 CAAATGAATCGTTGACACACCCCATCACAAAATACATCGCATGTCAGCT 290
QY 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLeuAla 40
|||||
Db 291 GATCTGGAGTAACCCAGCACCCTGGGTTTGTCTGGAGGGGTCTCGCGCCCTAGCG 350
QY 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyLys 60
|||||
Db 351 GCCTACTGCTTCAGTCGGTTGTGTGTGATGTGGTCAATCGAGCTGGGGGCAAG 410
QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluCys 80
|||||
Db 411 CGGCAATGTTCCAGACAAAGAGGTGTGTATCAACAATACGATGAGTGAAGAGTGC 470
QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGly-Ly 100
|||||
Db 471 TCACAAGCTGCCCATATATCGAAGCTCAGGTAAAGTCAAGTCAAGTCAA-GGAAAA 529
QY 100 sValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValTh 120
|||||
Db 530 AGTCCTGGATTGCTGCAGCAGGAGCCACCAACAAGCTGTCATTGAGCCCATAGTAA 589
QY 120 rThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
|||||
Db 590 TACCAACTGGCAAAAGCTTGAGGCGCTTTTGGCACAAGCAT 629

RESULT 9

US-09-899-046-31
; Sequence 31, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/899,046
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/362,455
FILING DATE:
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 401 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
MOLECULE TYPE: cdna
TOPOLOGY: linear
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: HD10-1-25
FEATURE:
NAME/KEY: CDS
LOCATION: 3..401
US-09-899-046-31

Alignment Scores:
Pred. No.: 4.57e-47 Length: 401
Score: 52.00 Matches: 87
Percent Similarity: 97.75% Conservative: 0
Best Local Similarity: 97.75% Mismatches: 1
Query Match: 39.10% Indels: 2
DB: 11 Gaps: 0

US-09-638-693-36 (1-133) x US-09-899-046-31 (1-401)

Qy 11 ThrLysTyrIleMetAlaCysMetSerAlaAspLeuGluValThrThrSerThrTrpVal 30
Db 33 ACAAATACATATTATGGCATGTCAGCTGATCTGGAAGTAACCCAGCACCTGGGTG 92
Qy 31 LeuLeuGlyGlyValLeuAlaAlaLeuAlaAlaTyrCysLeuSerValGlyCysValVal 50
Db 93 TTGCTTGGAGGGGTCCTCGCGGCCCTAGCGGCCCTACTGCTGTGTCAGTGGGTGTA 152
Qy 51 IleValGlyHisIleGluLeuGlyGlyLysProAlaIle-ValProAspLysGluValLe 70
Db 153 ATCGTGGGTATATCGAGCTGGGGGCAAGCCGCACT-CGTTCACACACAGGAGGTGT 211
Qy 70 uTyrGlnGlnTyrAspGluMetGluGluCysSerGlnAlaAlaProTyrIleGluGlnAl 90
Db 212 GTATCAACAGTACGATGAGATGGAGGAGTGTCTCGAAGCGCCCATACATCGAACAA 271
Qy 90 aGlnValIleAlaHisGlnPheLys 98
Db 272 TCAGGTAATAGCCACCACTTCAG 296

RESULT 10

US-09-899-046-33
Sequence 33, Application US/09899046
Publication No. US20030008274A1
GENERAL INFORMATION:

APPLICANT:
TITLE OF INVENTION: New sequences of hepatitis C virus
NUMBER OF SEQUENCES: 270
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/899,046
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/362,455
FILING DATE:
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:

LENGTH: 401 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: HD10-1-3
FEATURE:
NAME/KEY: CDS
LOCATION: 3..401
US-09-899-046-33

Alignment Scores:
Pred. No.: 4.57e-47 Length: 401
Score: 52.00 Matches: 87
Percent Similarity: 97.75% Conservative: 0
Best Local Similarity: 97.75% Mismatches: 1
Query Match: 39.10% Indels: 2
DB: 11 Gaps: 0

US-09-638-693-36 (1-133) x US-09-899-046-33 (1-401)

Qy 11 ThrLysTyrIleMetAlaCysMetSerAlaAspLeuGluValThrThrSerThrTrpVal 30
Db 33 ACAAATACATATTATGGCATGTCAGCTGATCTGGAAGTAACCCAGCACCTGGGTG 92
Qy 31 LeuLeuGlyGlyValLeuAlaAlaLeuAlaAlaTyrCysLeuSerValGlyCysValVal 50
Db 93 TTGCTTGGAGGGGTCCTCGCGGCCCTAGCGGCCCTACTGCTGTGTCAGTGGGTGTA 152
Qy 51 IleValGlyHisIleGluLeuGlyGlyLysProAlaIle-ValProAspLysGluValLe 70
Db 153 ATCGTGGGTATATCGAGCTGGGGGCAAGCCGCACT-CGTTCACACACAGGAGGTGT 211
Qy 70 uTyrGlnGlnTyrAspGluMetGluGluCysSerGlnAlaAlaProTyrIleGluGlnAl 90
Db 212 GTATCAACAGTACGATGAGATGGAGGAGTGTCTCGAAGCGCCCATACATCGAACAA 271
Qy 90 aGlnValIleAlaHisGlnPheLys 98
Db 272 TCAGGTAATAGCCACCACTTCAG 296

RESULT 11

US-09-878-281-31
Sequence 31, Application US/09878281
Publication No. US20030032005A1
GENERAL INFORMATION:

APPLICANT:
TITLE OF INVENTION: New sequences of hepatitis C virus
NUMBER OF SEQUENCES: 270
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/878,281
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/362,455
FILING DATE:
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 401 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
HYPOTHETICAL: NO
ANTI-SENSE: NO

IMMEDIATE SOURCE:
CLONE: HD10-1-25
FEATURE:
NAME/KEY: CDS
LOCATION: 3..401
US-09-878-281-31

Alignment Scores:
Pred. No.: 4,57e-47 Length: 401
Score: 52.00 Matches: 87
Percent Similarity: 97.75% Conservative: 0
Best Local Similarity: 97.75% Mismatches: 1
Query Match: 39.10% Indels: 2
DB: 11 Gaps: 0

US-09-638-693-36 (1-133) x US-09-878-281-31 (1-401)

QY 11 ThrLysTyrIleMetAlaCysMetSerAlaAspLeuGluValThrThrSerThrTrpVal 30
DB 33 ACAAAATACATTATGCGATGTCAGCTGATCTGGAAGTAACACACGACCTGGGTG 92
QY 31 LeuLeuGlyGlyValLeuAlaLeuAlaTyrCysLeuSerValGlyCysValVal 50
DB 93 TTGCTTGGAGGGGTCCTCGCGCCCTAGCGGCTACTGCTTGTGTCAGTCCGCTGGGTGTA 152
QY 51 IleValGlyHisIleGluLeuGlyGlyLysProAlaIle-ValProAspLysGluVal 70
DB 153 ATCGTGGGTTCATATCGAGCTGGGGGCAAGCCGGCACT-CGTTCACAGACAAGAGGTGT 211
QY 70 uTyrGlnGlnTyrAspGluMetGluCysSerGlnAlaAlaProTyrIleGluGlnAl 90
DB 212 GATCAACAGTACGATGAGTGGAGGAGTCTCGCAAGCGCCCATATCATCGAACAAGC 271
QY 90 aGlnValIleAlaHisGlnPheLys 98
DB 272 TCAGGTAATAGCCACCAAGTTCAG 296

RESULT 12

US-09-878-281-33
Sequence 33, Application US/09878281
Publication No. US20030032005A1
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: New sequences of hepatitis C virus
TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
NUMBER OF SEQUENCES: 270
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/878,281
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/362,455
FILING DATE:
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 401 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: HD10-1-3
FEATURE:
NAME/KEY: CDS
LOCATION: 3..401
US-09-878-281-33

Alignment Scores:
Pred. No.: 4,57e-47 Length: 401
Score: 52.00 Matches: 87
Percent Similarity: 97.75% Conservative: 0
Best Local Similarity: 97.75% Mismatches: 1
Query Match: 39.10% Indels: 2
DB: 11 Gaps: 0

US-09-638-693-36 (1-133) x US-09-878-281-33 (1-401)

QY 11 ThrLysTyrIleMetAlaCysMetSerAlaAspLeuGluValThrThrSerThrTrpVal 30
DB 33 ACAAAATACATTATGCGATGTCAGCTGATCTGGAAGTAACACACGACCTGGGTG 92
QY 31 LeuLeuGlyGlyValLeuAlaLeuAlaTyrCysLeuSerValGlyCysValVal 50
DB 93 TTGCTTGGAGGGGTCCTCGCGCCCTAGCGGCTACTGCTTGTGTCAGTCCGCTGGGTGTA 152
QY 51 IleValGlyHisIleGluLeuGlyGlyLysProAlaIle-ValProAspLysGluVal 70
DB 153 ATCGTGGGTTCATATCGAGCTGGGGGCAAGCCGGCACT-CGTTCACAGACAAGAGGTGT 211
QY 70 uTyrGlnGlnTyrAspGluMetGluCysSerGlnAlaAlaProTyrIleGluGlnAl 90
DB 212 GATCAACAGTACGATGAGTGGAGGAGTCTCGCAAGCGCCCATATCATCGAACAAGC 271
QY 90 aGlnValIleAlaHisGlnPheLys 98
DB 272 TCAGGTAATAGCCACCAAGTTCAG 296

RESULT 13

US-09-881-239-4
Sequence 4, Application US/09881239
Publication No. US20020192639A1
GENERAL INFORMATION:
APPLICANT: CHIEN, David Y.
APPLICANT: ARANGEL, Phillip
APPLICANT: TANDESKE, Laura
APPLICANT: GEORGE-NASCIEMENTO, Carlos
APPLICANT: COIT, Doris
APPLICANT: MEDINA-SELBY, Angelica
TITLE OF INVENTION: HCV ANTIGEN/ANTIBODY COMBINATION ASSAY
FILE REFERENCE: 2302-16073 / PP16073.003
CURRENT APPLICATION NUMBER: US/09/881,239
CURRENT FILING DATE: 2001-06-14
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 2499
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: MEFA 12
NAME/KEY: CDS
LOCATION: (1)..(2487)
US-09-881-239-4

Alignment Scores:
Pred. No.: 1.82e-34 Length: 2499
Score: 41.00 Matches: 41
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 30.83% Indels: 0
DB: 10 Gaps: 0

US-09-638-693-36 (1-133) x US-09-881-239-4 (1-2499)

QY 58 GlyGlyLysProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMet 77
DB 1330 GGGGGCAAGCCGGCAATCGTTCAGACAAAGAGGTGTGTATCAACAATACGATGAGATG 1389
QY 78 GluGluCysSerGlnAlaAlaProTyrIleGluGlnAlaAlaProTyrIleGlnPhe 97

```
Db 1390 GAAGAGTCTCACAGCTGCCCCATATATCGAACAAAGCTCAGGTAATAGTCACCGATTC 1449
Qy 98 Lys 98
   |||
Db 1450 AAG 1452

RESULT 14
US-09-881-654-3
; Sequence 3, Application US/09881654
; Patent No. US20020146685A1
; GENERAL INFORMATION:
; APPLICANT: CHIEN, David Y.
; APPLICANT: ARANGEL, Phillip
; APPLICANT: TANDESKE, Laura
; APPLICANT: GEORGE-NASCIEMENTO, Carlos
; APPLICANT: COLT, Doris
; APPLICANT: MEDINA-SELBY, Angelica
; TITLE OF INVENTION: IMMUNOASSAYS FOR ANTI-HCV ANTIBODIES
; FILE REFERENCE: 2302-17039 / PP17039.002
; CURRENT APPLICATION NUMBER: US/09/881.654
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: 60/212,082
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 60/280,811
; PRIOR FILING DATE: 2001-04-02
; PRIOR APPLICATION NUMBER: 60/280,867
; PRIOR FILING DATE: 2001-04-02
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 3297
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MEFA 7.1
; NAME/KEY: CDS
; LOCATION: (1)..(3297)
US-09-881-654-3

Alignment Scores:
Pred. No.: 2.34e-34 Length: 3297
Score: 41.00 Matches: 41
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 30.83% Indels: 0
DB: 10 Gaps: 0

US-09-638-693-36 (1-133) x US-09-881-654-3 (1-3297)

Qy 58 GlyGlyLysProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMet 77
   |||||
Db 2242 GGGGGCAAGCGGCAATCGTTCCAGACAAAGAGGTGTGTATCAACAATACGATGAGATG 2301

Qy 78 GluGluCysSerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPhe 97
   |||||
Db 2302 GAAGAGTCTCACAGCTGCCCATATATCGAACAAAGCTCAGGTAATAGTCACCGATTC 2361

Qy 98 Lys 98
   |||
Db 2362 AAG 2364

RESULT 15
US-09-899-046-29
; Sequence 29, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 287 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; CLONE: HC1153
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3...287
US-09-899-046-29

Alignment Scores:
Pred. No.: 1.38e-11 Length: 287
Score: 19.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.29% Indels: 0
DB: 11 Gaps: 0

US-09-638-693-36 (1-133) x US-09-899-046-29 (1-287)

Qy 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSer 19
   |||||
Db 231 CAAATGAAATCTGCTTGACACACACCCCATCAAAATACATCATGGCATGTCATGTC 287

Search completed: August 29, 2003, 14:28:10
Job time : 217 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 11:18:29 ; Search time 17 Seconds
(without alignments)
752.379 Million cell updates/sec

Title: US-09-638-693-36

Perfect score: 133

Sequence: 1 QNEICLTHPTIKYIMACMSA.....VIEPIVTTNWQKLEAFWHKH 133

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR_76:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	98	73.7	209	2	PC1306
2	52	39.1	142	2	PC1307
3	19	14.3	876	2	PC2219
4	19	14.3	3010	1	GNWVTC
5	19	14.3	3010	1	GNWVCJ
6	19	14.3	3010	1	GNWVTW
7	19	14.3	3014	1	JCS620
8	13	9.8	492	2	PS0326
9	13	9.8	716	2	QJ1366
10	13	9.8	3011	1	GNWVC3
11	13	9.8	3011	1	GNWVCH
12	13	9.8	3011	1	S40770
13	12	9.0	3010	1	A45573
14	12	9.0	3010	1	S18030
15	9	6.8	280	2	B84219
16	8	6.0	327	2	S61982
17	8	6.0	342	2	AD3450
18	8	6.0	363	1	E69889
19	8	6.0	399	2	E70598
20	8	6.0	420	2	C84201
21	7	5.3	80	2	S54909
22	7	5.3	96	2	H83065
23	7	5.3	107	2	G83545
24	7	5.3	125	2	S35629
25	7	5.3	128	2	B70647
26	7	5.3	159	2	D95864
27	7	5.3	172	2	B84260
28	7	5.3	174	2	C72737
29	7	5.3	184	2	A61196

30	7	5.3	188	2	AE3363	transcription regu
31	7	5.3	194	2	A54317	probable nonstruct
32	7	5.3	194	2	S06067	nonstructural prot
33	7	5.3	203	1	WMBP7B	gene 10 protein -
34	7	5.3	204	2	G95276	probable transcrip
35	7	5.3	217	2	B81067	conserved hypotet
36	7	5.3	217	2	G81801	probable lipoprote
37	7	5.3	222	2	B95207	aquaporin [impor
38	7	5.3	222	2	B98072	aquaporin %, wat
39	7	5.3	228	2	C83663	hypothetical prote
40	7	5.3	234	2	D96932	ABC transporter, A
41	7	5.3	235	2	E65082	hypothetical prote
42	7	5.3	245	2	AG2300	hypothetical prote
43	7	5.3	295	2	C71107	probable methionyl
44	7	5.3	295	2	JC5671	methionyl aminopep
45	7	5.3	303	2	AD3473	methyltransferase

ALIGNMENTS

RESULT 1

PC1306

genome polyprotein NS4a epitope containing region (isolate BR36-20) - hepatitis C vi.
C:Species: hepatitis C virus
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 17-Nov-2000
C:Accession: PC1306

R:Stuyver, L.; Van Arnhem, W.; Wyseur, A.; DeLeys, R.; Maertens, G.

Biochem. Biophys. Res. Commun. 192, 635-641, 1993

A:Title: Analysis of the putative E1 envelope and NS4a epitope regions of HCV type 3

A:Reference number: PC1300; MUID:93249436; PMID:7683463

A:Accession: PC1306

A:Molecule type: mRNA

A:Residues: 1-209 <STU>

A:Cross-references: DDBJ:DI4600; NID:g303584; PIDN:BAA03449.1; PID:g303585

A:Experimental source: blood

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: nonstructural protein; polyprotein

Query Match 73.7%; Score 98; DB 2; Length 209;

Best Local Similarity 100.0%; Pred. No. 7.7e-90;

Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QNEICLTHPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSGCVVIVGHIELGK 60

Db 77 QNEICLTHPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSGCVVIVGHIELGK 136

QY 61 PAIVDPKEVLYQQYDEMECSQAAPYIEQAQVIAHQFK 98

Db 137 PAIVDPKEVLYQQYDEMECSQAAPYIEQAQVIAHQFK 174

RESULT 2

PC1307

genome polyprotein NS4a epitope containing region (isolate HD10-1) - hepatitis C vir

C:Species: hepatitis C virus

C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 17-Nov-2000

C:Accession: PC1307

R:Stuyver, L.; Van Arnhem, W.; Wyseur, A.; DeLeys, R.; Maertens, G.

Biochem. Biophys. Res. Commun. 192, 635-641, 1993

A:Title: Analysis of the putative E1 envelope and NS4a epitope regions of HCV type 3

A:Reference number: PC1300; MUID:93249436; PMID:7683463

A:Accession: PC1307

A:Molecule type: mRNA

A:Residues: 1-142 <STU>

A:Cross-references: DDBJ:DI4602

A:Experimental source: blood

C:Superfamily: hepatitis C virus genome polyprotein

C:Keywords: polyprotein

Query Match 39.1%; Score 52; DB 2; Length 142;

Best Local Similarity 100.0%; Pred. No. 3.6e-44;

Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

A;Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
 A;Reference number: A40244; MUID:92230206; PMID:1314449
 A;Accession: A40244
 A;Molecule type: genomic RNA
 A;Residues: 1-3010 <CHE>

A;Cross-references: GB:M84754
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural
 F;1-115/Product: capsid protein C #status predicted <CPC>
 F;116-191/Product: major envelope protein M #status predicted <EPM>
 F;192-389/Product: major envelope protein E #status predicted <MEE>
 F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F;1007-1615/Product: hepatitis C virus genome polyprotein
 F;1230-1237/Region: nucleotide-binding motif A (P-loop)
 F;1312-1317/Region: nucleotide-binding motif B
 F;1316-1319/Region: DEXH motif
 F;1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
 F;1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
 F;2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F;196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,207

Query Match 14.3%; Score 19; DB 1; Length 3010;

Best Local Similarity 100.0%; Pred. No. 3.3e-10;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 LTHPTIKYIMACMSADLEV 24

|||||

Db 1637 LTHPTIKYIMACMSADLEV 1655

RESULT 7

JC5620

genome polyprotein - hepatitis C virus (isolate EUH1480)
 N;Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
 C;Superfamily: hepatitis C virus genome polyprotein
 C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin

F;116-191/Product: capsid protein C #status predicted <CPC>
 F;192-389/Product: envelope protein M #status predicted <EPM>
 F;390-729/Product: major envelope protein E #status predicted <MEE>
 F;731-1007/Product: nonstructural protein NS1 #status predicted <NS1>
 F;1008-1616/Product: hepatitis C virus genome polyprotein
 F;1231-1238/Region: nucleotide-binding motif A (P-loop)
 F;1313-1318/Region: nucleotide-binding motif B
 F;1317-1320/Region: DEXH motif
 F;1617-1863/Product: nonstructural protein NS4a #status predicted <NS4a>
 F;1864-2014/Product: nonstructural protein NS4b #status predicted <NS4b>
 F;2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
 F;2210-2249/Region: interferon sensitivity determining #status predicted

C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001

C;Accession: JC5620

R;Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.

Biochem. Biophys. Res. Commun. 236, 44-49, 1997

A;Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant

A;Reference number: JC5620; MUID:97366593; PMID:9223423

A;Accession: JC5620

A;Molecule type: mRNA

A;Residues: 1-3014 <CHA>

A;Cross-references: GB:Y13184

A;Experimental source: genotype 5a, which predominates in South Africa

A;Note: the translation of the nucleotide sequence is not complete in this paper

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin

F;116-191/Product: capsid protein C #status predicted <CPC>
 F;192-389/Product: envelope protein M #status predicted <EPM>
 F;390-729/Product: major envelope protein E #status predicted <MEE>
 F;731-1007/Product: nonstructural protein NS1 #status predicted <NS1>
 F;1008-1616/Product: hepatitis C virus genome polyprotein
 F;1231-1238/Region: nucleotide-binding motif A (P-loop)
 F;1313-1318/Region: nucleotide-binding motif B
 F;1317-1320/Region: DEXH motif
 F;1617-1863/Product: nonstructural protein NS4a #status predicted <NS4a>
 F;1864-2014/Product: nonstructural protein NS4b #status predicted <NS4b>
 F;2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
 F;2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 14.3%; Score 19; DB 1; Length 3014;

Best Local Similarity 100.0%; Pred. No. 3.3e-10;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 LTHPTIKYIMACMSADLEV 24

|||||

Db 1638 LTHPTIKYIMACMSADLEV 1656

RESULT 8

PS0326

polyprotein - hepatitis C virus (isolate Fla) (fragments)

C;Species: hepatitis C virus

C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000

C;Accession: PS0326

R;Li, J.S.; Tong, S.P.; Vitvitski, L.; Lepot, D.; Trepo, C.

Gene 105, 167-172, 1991

A;Title: Two French genotypes of hepatitis C virus: homology of the predominant geno

A;Reference number: PS0326; MUID:92039028; PMID:1718820

A;Accession: PS0326

A;Molecule type: genomic RNA

A;Residues: 1-492 <LIJ>

A;Cross-references: GB:M60220

A;Note: this sequence corresponds to nonstructural protein NS3 region

A;Note: translation of the nucleotide sequence is not complete

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: polyprotein

Query Match 9.8%; Score 13; DB 2; Length 492;

Best Local Similarity 100.0%; Pred. No. 6.8e-05;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 GGVLAALAAAYCLS 45

|||||

Db 215 GGVLAALAAAYCLS 227

RESULT 9

JQ1366

polyprotein - hepatitis C virus (French isolate) (fragments)

C;Species: hepatitis C virus

C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000

C;Accession: JQ1366

R;Kremsdorf, D.; Porchon, C.; Kim, J.P.; Reyes, G.R.; Brechot, C.

J. Gen. Virol. 72, 2557-2561, 1991

A;Title: Partial nucleotide sequence analysis of a French hepatitis C virus: implica

A;Reference number: JQ1366; MUID:92013977; PMID:1655961

A;Accession: JQ1366

A;Molecule type: genomic RNA

A;Residues: 1-716 <KRE>

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: glycoprotein; polyprotein

F;84,90,97,115,143,199,223,243,290,312/Binding site: carbohydrate (Asn) (covalent) #

Query Match 9.8%; Score 13; DB 2; Length 716;

Best Local Similarity 100.0%; Pred. No. 9.4e-05;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 GGVLAALAAAYCLS 45

|||||

Db 613 GGVLAALAAAYCLS 625

RESULT 10

GNVVC3

genome polyprotein - hepatitis C virus (strain HCV-1)

N;Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein

protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C;Species: hepatitis C virus

C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 19-Jan-2001

C;Accession: A39166; PQ0403; PQ0404

R;Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.;

Proc. Natl. Acad. Sci. U.S.A. 88, 2451-2455, 1991

A;Title: Genetic organization and diversity of the hepatitis C virus.

A;Reference number: A39166; MUID:91172826; PMID:1848704

A;Accession: A39166

A;Molecule type: mRNA

A;Residues: 1-3011 <CHO>

A;Cross-references: GB:M62321; NID:g329873; PIDN:AAA5676.1; PID:g329874

R;Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peuthner, J.F.; Follett, E.; Yap,

J. Gen. Virol. 73, 1131-1141, 1992
 A:Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e
 A:Reference number: PQ0393; MUID:92268871; PMID:1316939
 A:Accession: PQ0403
 A:Molecule type: genomic RNA
 A:Residues: 1577-1633 <CHA>
 A:Cross-references: DBJ:DI0128
 A:Experimental source: Isolates E-b16
 A:Accession: PQ0404
 A>Status: preliminary
 A:Molecule type: genomic RNA
 A:Residues: 1577-1633 <CH2>
 A:Experimental source: Isolates E-b17
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: capsid protein C #status predicted <CPC>
 F:192-389/Product: major envelope protein E #status predicted <MEE>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus genome polyprotein
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22

Query Match 9.8%; Score 13; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 0.00031;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 GGVLAALAAAYCLS 45
 |||||
 Db 1664 GGVLAALAAAYCLS 1676

RESULT 11
 GNMVCH
 genome polyprotein - hepatitis C virus (strain H)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 A:Note: host homo sapiens (man)
 C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
 C:Accession: A36814; A41546
 R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
 submitted to GenBank, July 1992
 A:Description: Genomic structure of the human prototype strain H of hepatitis C virus: c
 A:Reference number: A36814
 A:Accession: A36814
 A:Molecule type: genomic RNA
 A:Residues: 1-3011 <INC>
 A:Cross-references: GB:M67463; NID:g329737; PIDN:AAA45534.1; PID:g329738
 R:Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
 Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991
 A:Title: Genomic structure of the human prototype strain H of hepatitis C virus: compar
 A:Reference number: A41546; MUID:92052256; PMID:1658800
 A:Contents: annotation
 A:Note: neither amino acid nor nucleotide sequence is given
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: capsid protein C #status predicted <CPC>
 F:192-389/Product: major envelope protein E #status predicted <MEE>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus genome polyprotein
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240

Query Match 9.8%; Score 13; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 0.00031;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 GGVLAALAAAYCLS 45
 |||||
 Db 1664 GGVLAALAAAYCLS 1676

RESULT 12
 S40770
 genome polyprotein - hepatitis C virus
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
 C:Accession: S40770; PC1285
 R:Okamoto, H.
 submitted to the EMBL Data Library, March 1992
 A:Reference number: S40770
 A:Accession: S40770
 A:Molecule type: genomic RNA
 A:Residues: 1-3011 <OKA>
 A:Cross-references: EMBL:D10749; NID:g221586; PIDN:BAA01582.1; PID:g221587
 R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Totsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tsu
 Jpn. J. Exp. Med. 60, 167-177, 1990
 A:Title: The 5'-terminal sequence of the hepatitis C virus genome.
 A:Reference number: PC1284; MUID:91013116; PMID:2170712
 A:Accession: PC1285
 A:Molecule type: genomic RNA
 A:Residues: 1-513 <OK2>
 A:Cross-references: GB:D00831; NID:g221511; PIDN:BAA00705.1; PID:g221512
 A:Experimental source: isolate HC-J1
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; se
 F:2-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: capsid protein C #status predicted <CPC>
 F:192-389/Product: major envelope protein E #status predicted <MEE>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: hepatitis C virus genome polyprotein
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 9.8%; Score 13; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 0.00031;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 GGVLAALAAAYCLS 45
 |||||
 Db 1664 GGVLAALAAAYCLS 1676

RESULT 13
 A45573
 genome polyprotein - hepatitis C virus (strain JT)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
 protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
 C:Accession: A45573
 R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikat
 Virus Res. 23, 39-53, 1992
 A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier
 A:Reference number: A45573; MUID:92295714; PMID:1318627

A:Accession: A45573
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3010 <TAN>
A:Cross-references: GB:D11168; GB:D01171; NID:G221612; PIDN:BAA01943.1; PID:G221613
A:Experimental source: HCV-JT
A:Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBIP:106207)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus polyprotein #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>
F:1616-1862/Product: nonstructural protein NS4b #status predicted <NS4b>
F:1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 9.0%; Score 12; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 0.0031;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 GGVLAALAAAYCL 44
|||||
Db 1664 GGVLAALAAAYCL 1675

RESULT 14
S18030
genome polyprotein - hepatitis C virus (isolate Jk1)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (isolate Jk1)
C:Species: hepatitis C virus
A:Variety: isolate Jk1
C:Accession: S18030; S33570; A48332; S18029
R:Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.
submitted to the EMBL data library, September 1991
A:Description: A whole genome of hepatitis C virus cDNA was isolated from a single patie
A:Reference number: S18028
A:Accession: S18030
A:Molecule type: genomic RNA
A:Residues: 1-3010 <HON>
A:Cross-references: EMBL:X61596; NID:G59478; PIDN:CAA43793.1; PID:G59479
A:Experimental source: isolate Jk1 from an individual
R:Honda, M.; Kaneko, S.; Unoura, M.; Kobayashi, K.; Murakami, S.
Arch. Virol. 128, 163-169, 1993
A:Title: Sequence analysis of putative structural regions of hepatitis C virus isolated
A:Reference number: A48332; MUID:93119270; PMID:8380322
A:Accession: S33570
A:Molecule type: genomic RNA
A:Residues: 1-547,'T','V',623-624,'S',626-652,'DL',655-761,'T',763-782 <HOW>
A:Cross-references: EMBL:X61591
A:Note: this sequence is inconsistent with the nucleotide translation
A:Note: the authors translated the codon AGG for residue 43 as Pro, TGG for residue 320
as Trp, and TTC for residue 771 as Ser
A:Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBIP:121748)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus polyprotein #status predicted <NS3>
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>

F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196-209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate

Query Match 9.0%; Score 12; DB 1; Length 3010;
Best Local Similarity 100.0%; Pred. No. 0.0031;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 GGVLAALAAAYCL 44
|||||
Db 1664 GGVLAALAAAYCL 1675

RESULT 15
B84219
4-hydroxybenzoate octaprenyltransferase [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 01-Mar-2002
C:Accession: B84219
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jn
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: B84219
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-280 <STO>
A:Cross-references: GB:AE004437; NID:G10580202; PIDN:AAG19118.1; GSPDB:GN00138
C:Genetics:
A:Gene: rhoA
C:Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ0279

Query Match 6.8%; Score 9; DB 2; Length 280;
Best Local Similarity 100.0%; Pred. No. 0.41;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 GGVLAALAA 41
|||||
Db 161 GGVLAALAA 169

Search completed: August 29, 2003, 11:24:01
Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 11:18:59 ; Search time 11 seconds
(without alignments)
568.596 Million cell updates/sec

Title: US-09-638-693-36
Perfect score: 133
Sequence: 1 QNEICLTHPIKYNACMSA.....VIEPIVTNWKLEAFWHKH 133

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	14.3	3010	1 POLG_HCVBK	P26663 h genome po
2	19	14.3	3010	1 POLG_HCVJA	P26662 h genome po
3	19	14.3	3010	1 POLG_HCVTW	P29846 h genome po
4	13	9.8	3011	1 POLG_HCV1	P26664 h genome po
5	13	9.8	3011	1 POLG_HCVH	P27958 h genome po
6	12	9.0	3010	1 POLG_HCVJT	Q00269 h genome po
7	7	5.3	80	1 GCHL_EUGGR	P51597 euglena gra
8	7	5.3	128	1 NUOA_MYCTU	P95181 mycobacteri
9	7	5.3	203	1 VP10_BPPRD	P28732 bacterioph
10	7	5.3	228	1 ISPD_BACHD	Q9kf88 bacillus ha
11	7	5.3	235	1 YOGA_ECOLI	Q46831 escherichia
12	7	5.3	253	1 ISPD_RALSO	Q8xyw3 ralstonia s
13	7	5.3	276	1 SNAB_STRPR	P54993 streptomyce
14	7	5.3	295	1 AMPM_PYRFU	P56218 pyrococcus
15	7	5.3	295	1 AMPM_PYRHO	O58362 pyrococcus
16	7	5.3	306	1 ISPE_MYCTU	O05596 mycobacteri
17	7	5.3	344	1 Y401_HUMAN	Q43151 homo sapien
18	7	5.3	351	1 PYRD_XYLF	Q9pae7 xyella fas
19	7	5.3	366	1 Y121_SYNY3	P73714 synechocyst
20	7	5.3	419	1 MURA_ACIGB	P33986 acinetobact
21	7	5.3	419	1 YC58_MYCTU	Q11060 mycobacteri
22	7	5.3	427	1 DAMX_ECOLI	P11557 escherichia
23	7	5.3	450	1 AROA_MYCTU	P22487 mycobacteri
24	7	5.3	484	1 MURE_RHILO	Q98ka8 rhizobium l
25	7	5.3	538	1 TNSE_ECOLI	P05845 escherichia
26	7	5.3	609	1 Y4PA_RHISN	P55610 rhizobium s
27	7	5.3	797	1 CTPE_MYCTU	O08365 mycobacteri
28	7	5.3	815	1 NAHL_HUMAN	P19634 homo sapien
29	7	5.3	816	1 NAHL_RABIT	P23791 oryctolagus
30	7	5.3	817	1 NAHL_BOVIN	Q28036 bos taurus
31	7	5.3	818	1 NAHL_PIG	Q48762 sus scrofa
32	7	5.3	820	1 NAHL_MOUSE	Q61165 mus musculu
33	7	5.3	820	1 NAHL_RAT	P26431 rattus norv

RESULT 1				
ID	POLG_HCVBK	STANDARD;	PRT;	3010 AA.
AC	P26663;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);			
DE	Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2			
DE	(GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)			
DE	(EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)			
DE	(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein			
DE	NS4B (P66); Nonstructural protein NS5A (P56); Nonstructural protein			
OS	Hepatitis C virus (isolate BK) (HCV).			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;			
OC	Hepacivirus.			
OX	NCBI_TaxID=11105;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91140698; PubMed=1847440;			
RA	Takamizawa A., Mori C., Fuke I., Manabe S., Murakami S., Fujita J.,			
RA	Onishi E., Andoh T., Yoshida I., Okayama H.;			
RT	"Structure and organization of the hepatitis C virus genome isolated			
RT	from human carriers.";			
RL	J. Virol. 65:1105-1113(1991).			
RN	[2]			
RP	SEQUENCE OF 1487-1500.			
RX	MEDLINE=96235224; PubMed=8647104;			
RA	Borowski P., Heiland M., Oehlmann K., Becker B., Kornetevy L.;			
RT	"Non-structural protein 3 of hepatitis C virus inhibits			
RT	phosphorylation mediated by cAMP-dependent protein kinase.";			
RL	Eur. J. Biochem. 237:611-618(1996).			
RN	[3]			
RP	X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.			
RX	MEDLINE=97015088; PubMed=8861916;			
RA	Love R.A., Parge H.E., Wickersham J.A., Hostomsky Z., Habuka N.,			
RA	Moomaw E.W., Adachi T., Hostomsky Z.;			
RT	"The crystal structure of hepatitis C virus NS3 proteinase reveals a			
RT	trypsin-like fold and a structural zinc binding site.";			
RL	Cell 87:331-342(1996).			
RN	[4]			
RP	X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.			
RX	MEDLINE=98227846; PubMed=9568891;			
RA	Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,			
RA	Steinkuehler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;			
RT	"Complex of NS3 protease and NS4A peptide of BK strain hepatitis C			
RT	virus: a 2.2-A resolution structure in a hexagonal crystal form.";			
RL	Protein Sci. 7:837-847(1998).			
CC	!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE			
CC	HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.			
CC	NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.			
CC	!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral			
CC	precursor polyprotein, commonly with Asp or Glu in the P6			
CC	position, Cys or Thr in P1 and Ser or Ala in P1'.			
CC	!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +			

P11718 leishmania
P12522 leishmania
P57680 mus musculu
P33479 pseudorabie
P11675 pseudorabie
P26660 h genome po
P03791 bacterioph
O88005 bordetella
P08379 locusta mig
P32691 escherichia
P51598 mucuna hass
P34613 caenorhabdi

ALIGNMENTS

```

CC [RNA](N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M58335; AAA72945.1; .
CC PIR: A38465; GNVVTC.
CC PDB: 1AIQ; 25-MAR-98.
CC PDB: 1JXP; 14-JAN-98.
CC PDB: 1NS3; 08-APR-98.
CC PDB: 1C2P; 15-NOV-00.
CC PDB: 1CSJ; 08-NOV-99.
CC PDB: 1GX5; 09-APR-02.
CC PDB: 1GX6; 10-APR-02.
CC PDB: 1QUV; 26-JUN-00.
CC PDB: 80HM; 20-APR-99.
CC MEROPS: S29.001; .
CC MEROPS: U39.001; .
CC InterPro: IPR001410; DEAD.
CC InterPro: IPR002522; HCV_capsid.
CC InterPro: IPR002521; HCV_core.
CC InterPro: IPR002519; HCV_env.
CC InterPro: IPR002531; HCV_NS1.
CC InterPro: IPR002518; HCV_NS2.
CC InterPro: IPR004109; HCV_NS3.
CC InterPro: IPR000745; HCV_NS4a.
CC InterPro: IPR001490; HCV_NS4b.
CC InterPro: IPR002868; HCV_NS5a.
CC InterPro: IPR002166; HCV_RdRP.
CC InterPro: IPR007095; RNA_pol_DS_Ps.
CC InterPro: IPR007094; RNA_pol_PSVir.
CC Pfam: PF01543; HCV_capsid; 1.
CC Pfam: PF01542; HCV_core; 1.
CC Pfam: PF01539; HCV_env; 1.
CC Pfam: PF01560; HCV_NS1; 1.
CC Pfam: PF01538; HCV_NS2; 1.
CC Pfam: PF02907; HCV_NS3; 1.
CC Pfam: PF01006; HCV_NS4a; 1.
CC Pfam: PF01001; HCV_NS4b; 1.
CC Pfam: PF01506; HCV_NS5a; 1.
CC Pfam: PF00998; Viral_RdRP; 1.
CC ProDom: PD186062; HCV_NS1; 1.
CC SMART: SM00487; DEXDC; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
KW 3D-structure.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CELLULAR AMINOPEPTIDASE.
FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 116 191 MATRIX PROTEIN E (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM.
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM.
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM.
FT NP_BIND 1230 1237 ATP (POTENTIAL).

```

```

FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
FT STRAND 1031 1035 N-LINKED (GLCNAC. .) (POTENTIAL).
FT HELIX 1039 1047 N-LINKED (GLCNAC. .) (POTENTIAL).
FT STRAND 1050 1050 N-LINKED (GLCNAC. .) (POTENTIAL).
FT STRAND 1059 1063 N-LINKED (GLCNAC. .) (POTENTIAL).
FT STRAND 1074 1074 N-LINKED (GLCNAC. .) (POTENTIAL).
FT TURN 1075 1076 N-LINKED (GLCNAC. .) (POTENTIAL).
FT STRAND 1077 1081 N-LINKED (GLCNAC. .) (POTENTIAL).
FT TURN 1082 1085 N-LINKED (GLCNAC. .) (POTENTIAL).
FT STRAND 1086 1087 N-LINKED (GLCNAC. .) (POTENTIAL).
FT TURN 1090 1092 N-LINKED (GLCNAC. .) (POTENTIAL).
FT TURN 1093 1094 N-LINKED (GLCNAC. .) (POTENTIAL).
FT STRAND 1095 1097 N-LINKED (GLCNAC. .) (POTENTIAL).
FT STRAND 1101 1103 N-LINKED (GLCNAC. .) (POTENTIAL).
FT TURN 1104 1107 N-LINKED (GLCNAC. .) (POTENTIAL).
FT STRAND 1108 1112 N-LINKED (GLCNAC. .) (POTENTIAL).
FT STRAND 1120 1120 N-LINKED (GLCNAC. .) (POTENTIAL).
FT STRAND 1122 1122 N-LINKED (GLCNAC. .) (POTENTIAL).
FT STRAND 1133 1133 N-LINKED (GLCNAC. .) (POTENTIAL).
FT TURN 1135 1136 N-LINKED (GLCNAC. .) (POTENTIAL).
FT STRAND 1139 1144 N-LINKED (GLCNAC. .) (POTENTIAL).
FT STRAND 1149 1157 N-LINKED (GLCNAC. .) (POTENTIAL).
FT TURN 1158 1161 N-LINKED (GLCNAC. .) (POTENTIAL).
FT TURN 1162 1163 N-LINKED (GLCNAC. .) (POTENTIAL).
FT TURN 1165 1166 N-LINKED (GLCNAC. .) (POTENTIAL).
FT STRAND 1168 1171 N-LINKED (GLCNAC. .) (POTENTIAL).
FT TURN 1172 1174 N-LINKED (GLCNAC. .) (POTENTIAL).
FT STRAND 1175 1186 N-LINKED (GLCNAC. .) (POTENTIAL).
FT TURN 1187 1188 N-LINKED (GLCNAC. .) (POTENTIAL).
FT STRAND 1189 1197 N-LINKED (GLCNAC. .) (POTENTIAL).
FT HELIX 1198 1202 N-LINKED (GLCNAC. .) (POTENTIAL).
FT TURN 1203 1204 N-LINKED (GLCNAC. .) (POTENTIAL).
FT STRAND 1680 1688 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECCFDFD9C CRC64;

```

Query Match 14.3%; Score 19; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 1.4e-10;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 LTHPTKYIMACMSADLEY 24
 |||||
 Db 1637 LTHPTKYIMACMSADLEY 1655

RESULT 2

POLG_HCVJA STANDARD; PRT; 3010 AA.
 ID POLG_HCVJA AC P26662;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2

CC [RNA] (N).

CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A

CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF

CC PROTEIN C AND MRNA.

CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC -----

CC EMBL; M62321; AAM45676.1; -

CC PIR; A39166; GNWVC3.

CC PDB; 1A1V; 16-FEB-99.

CC PDB; 1HEI; 25-NOV-98.

CC MEROPS; S29.001; -

CC MEROPS; U39.001; -

CC InterPro; IPR001410; DEAD.

CC InterPro; IPR002522; HCV_capsid.

CC InterPro; IPR002521; HCV_core.

CC InterPro; IPR002519; HCV_env.

CC InterPro; IPR002531; HCV_NS1.

CC InterPro; IPR002518; HCV_NS2.

CC InterPro; IPR004109; HCV_NS3.

CC InterPro; IPR000745; HCV_NS4a.

CC InterPro; IPR001490; HCV_NS4b.

CC InterPro; IPR002868; HCV_NS5a.

CC InterPro; IPR002166; HCV_RdRp.

CC InterPro; IPR001650; Helicase_C.

CC InterPro; IPR007095; RNA_pol_DS_PS.

CC InterPro; IPR007094; RNA_pol_PSVir.

CC Pfam; PF01543; HCV_core; 1.

CC Pfam; PF01542; HCV_core; 1.

CC Pfam; PF01539; HCV_env; 1.

CC Pfam; PF01560; HCV_NS1; 1.

CC Pfam; PF01538; HCV_NS2; 1.

CC Pfam; PF02907; HCV_NS3; 1.

CC Pfam; PF01006; HCV_NS4a; 1.

CC Pfam; PF01001; HCV_NS4b; 1.

CC Pfam; PF01506; HCV_NS5a; 1.

CC Pfam; PF00271; helicase_C; 1.

CC Pfam; PF00998; Viral_RdRp; 1.

CC ProDom; PD186062; HCV_NS1; 1.

CC SMART; SM00487; DEXDc; 1.

CC Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;

CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;

CC Transmembrane; Nonstructural protein; Hydrolase; Serine protease;

CC 3D-structure.

CC INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE

CC CELLULAR AMINOPEPTIDASE.

CC CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).

CC CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).

CC CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).

CC CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).

CC CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).

CC CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).

CC CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).

CC CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).

CC CHAIN 2014 3011 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).

CC CHAIN 347 369 POTENTIAL.

CC TRANSMEM 347 369 CHARGE RELAY SYSTEM (BY SIMILARITY).

CC ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).

CC ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).

CC ACT_SITE 1165 1165 ATP (POTENTIAL).

CC NP_BIND 1230 1237 DESCH BOX.

CC SITE 1316 1319 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).

CC CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).

SO SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Query Match 9.8%; Score 13; DB 1; Length 3011;

Best Local Similarity 100.0%; Pred. No. 0.00013;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 GGVLAALAAAYCLS 45

DB 1664 GGVLAALAAAYCLS 1676

RESULT 5

POIG_HCVH STANDARD; PRT; 3011 AA.

ID POLG_HCVH

AC P27958;

DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);

DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2

DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)

DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)

DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein

DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein

DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

OS Hepatitis C virus (isolate H) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI_TaxID=11108;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92052256; PubMed=1658800;

RA Inchauspe G., Zebadee S., Lee D.H.H., Sugitani M., Nasoff M.,

RA Prince A.M.;

RT "Genomic structure of the human prototype strain H of hepatitis C

RT virus: comparison with American and Japanese isolates.";

RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).

RN [2]

RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.

RX MEDLINE=97331322; PubMed=9187654;

RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;

RT "Structure of the hepatitis C virus RNA helicase domain.";

RL Nat. Struct. Biol. 4:463-467(1997).

RN [3]

RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.

RX MEDLINE=98154321; PubMed=9493270;

RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,

RA Murcko M.A., Lin C., Caron P.R.;

RT "Hepatitis C virus NS3 RNA helicase domain with a bound

RT oligonucleotide: the crystal structure provides insights into the mode

RT of unwinding.";

RL Structure 6:89-100(1998).

CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.

CC -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF

CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.

CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE

CC ACTIVATION OF NS3.

FT TURN 1598 1598
 FT HELIX 1606 1611
 FT TURN 1614 1618
 FT STRAND 1622 1623
 FT STRAND 1627 1627
 FT STRAND 1635 1636
 FT HELIX 1640 1652
 SQ SEQUENCE 3011 AA; 327142 MW; 772CBB29CCD94753 CRC64;

Query Match 9.8%; Score 13; DB 1; Length 3011;
 Best Local Similarity 100.0%; Pred. No. 0.00013;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 GGVLAALAAAYCLIS 45
 |||||
 DB 1664 GGVLAALAAAYCLIS 1676

RESULT 6

POLG.HCVJT STANDARD; PRT; 3010 AA.
 ID POLG.HCVJT
 AC Q00769;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (Isolate HC-JT) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 CC NCBI_TaxID=31642;
 RX SEQUENCE FROM N.A.
 RA MEDLINE=92295714; PubMed=1318627;
 RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
 RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
 RT "Molecular cloning of hepatitis C virus genome from a single Japanese
 RT carrier: sequence variation within the same individual and among
 RT infected individuals.";
 RL Virus Res. 23:39-53(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC [RNA](N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; D11168; BAA01943.1; -
 CC PIR; A45573; A45573.
 CC PDB; 1AIQ; 25-MAR-98.
 CC PDB; 1JXP; 14-JAN-98.
 CC MEROPS; S29.001; -
 CC MEROPS; U39.001; -

DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR002522; HCV_capsid.
 DR InterPro; IPR002521; HCV_core.
 DR InterPro; IPR002519; HCV_env.
 DR InterPro; IPR002531; HCV_NS1.
 DR InterPro; IPR002518; HCV_NS2.
 DR InterPro; IPR004109; HCV_NS3.
 DR InterPro; IPR000745; HCV_NS4a.
 DR InterPro; IPR001490; HCV_NS4b.
 DR InterPro; IPR002868; HCV_NS5a.
 DR InterPro; IPR002166; HCV_RDRP.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR Pfam; PF01543; HCV_capsid; 1.
 DR Pfam; PF01542; HCV_core; 1.
 DR Pfam; PF01539; HCV_env; 1.
 DR Pfam; PF01560; HCV_NS1; 1.
 DR Pfam; PF01538; HCV_NS2; 1.
 DR Pfam; PF02907; HCV_NS3; 1.
 DR Pfam; PF01006; HCV_NS4a; 1.
 DR Pfam; PF01001; HCV_NS4b; 1.
 DR Pfam; PF01506; HCV_NS5a; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR Pfam; PF00998; Viral_RDRP; 1.
 DR ProDom; PD186062; HCV_NS1; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT TRANSMEM 347 369 POTENTIAL.
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 3010 AA; 326573 MW; 94A1C77435D642B CRC64;

Query Match 9.0%; Score 12; DB 1; Length 3010;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 33 GGVLAALAAAYCL 44
    |||||
Db 1664 GGVLAALAAAYCL 1675

RESULT 7
GCHI_EUGGR STANDARD; PRT; 80 AA.
AC P51597;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE GTP cyclohydrolase I (EC 3.5.4.16) (GTP-CH-I) (Fragment).
OS Euglena gracilis.
OC Eukaryota; Euglenozoa; Euglenida; Euglenales; Euglena.
OX NCBI_TaxID=3039;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z / ATCC 12894;
RX MEDLINE=95352066; PubMed=7542887;
RA Maier J., Witter K., Guetlich M., Ziegler I., Werner T., Ninnemann H.;
RT "Homology cloning of GTP-cyclohydrolase I from various unrelated
RT eukaryotes by reverse-transcription polymerase chain reaction using a
RT general set of degenerate primers".
RL Biochem. Biophys. Res. Commun. 212:705-711(1995).
CC -1- CATALYTIC ACTIVITY: GTP + 2 H(2)O = formate + 2-amino-4-hydroxy-6-
CC (erythro-1,2,3-trihydroxypropyl)dihydropteridine triphosphate.
CC -1- ENZYME REGULATION: GTP SHOWS A POSITIVE ALLOSTERIC EFFECT, AND
CC TETRAHYDROBIPTERIN INHIBITS THE ENZYME ACTIVITY (BY SIMILARITY).
CC -1- PATHWAY: Tetrahydrobiopterin biosynthesis; first step.
CC -1- SUBUNIT: Homopolymer (By similarity).
CC -1- SIMILARITY: BELONGS TO THE GTP CYCLOHYDROLASE I FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z49757; CAA89827.1; -.
DR PIR; S54909; S54909.
DR HSP; P27511; IGTP.
DR InterPro; IPR001474; GTP_cyclohydrol.
DR Pfam; PF01227; GTP_cyclohydrol; 1.
DR ProDom; PD003330; GTP_cyclohydrol; 1.
DR PROSITE; PS00859; GTP_CYCLOHYDROL_1_1; PARTIAL.
DR PROSITE; PS00860; GTP_CYCLOHYDROL_1_2; 1.
KW Tetrahydrobiopterin biosynthesis; Hydrolase; Allosteric enzyme.
FT NON_TER 1 1
FT DISULFID 4 75 BY SIMILARITY.
FT NON_TER 80 80
SQ SEQUENCE 80 AA; 9097 MW; 8C46A15D01A7B8C7 CRC64;

Query Match 5.3%; Score 7; DB 1; Length 80;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 98 KGVVLGL 104
    |||||
Db 22 KGVVLGL 28

RESULT 8
NUOA_MYCTU STANDARD; PRT; 128 AA.
AC P55181;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE NADH-quinone oxidoreductase chain A (EC 1.6.99.5) (NADH dehydrogenase
DE I, chain A) (NDH-1, chain A).

```

```

GN NUOA OR RV3145 OR MT3233 OR MTCY03A2.13C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: NDH-1 shuttles electrons from NADH, via FMN and iron-
CC sulfur (Fe-S) centers, to quinones in the respiratory chain. The
CC immediate electron acceptor for the enzyme in this species is
CC believed to be menaquinone. Couples the redox reaction to proton
CC translocation (for every two electrons transferred, four hydrogen
CC ions are translocated across the cytoplasmic membrane), and thus
CC conserves the redox energy in a proton gradient (By similarity).
CC -1- CATALYTIC ACTIVITY: NADH + quinone = NAD(+) + quinol.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 3 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z83867; CAB06271.1; -.
DR EMBL; AE007138; AAK47572.1; -.
DR PIR; B70647; B70647.
DR TIGR; MT3233; -.
DR TubercuList; RV3145; -.
DR InterPro; IPR000440; Oxidored_q4.
DR Pfam; PF00507; oxidored_q4; 1.
DR KW Oxidoreductase; NAD; Quinone; Transmembrane; Complete proteome.
FT TRANSMEM 5 25 POTENTIAL.
FT TRANSMEM 72 92 POTENTIAL.
FT TRANSMEM 100 120 POTENTIAL.
SQ SEQUENCE 128 AA; 13975 MW; DDF4FD3F77A8149A CRC64;

Query Match 5.3%; Score 7; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 35 VLAALAA 41
    |||||
Db 9 VLAALAA 15

RESULT 9

```



```

VP10_BPPRD
ID VP10_BPPRD STANDARD; PRT; 203 AA.
AC P28732;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Protein P10.
GN X.
OS Bacteriophage PRD1.
OC Viruses; dsDNA viruses, no RNA stage; Tectiviridae; Tectivirus.
OX NCBI_TaxID=10658;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91306449; PubMed=1853567;
RA Bamford J.K.H., Haeninen A.-L., Pakula T.M., Ojala P.M.,
RA Kalkkinen N., Frilander M., Bamford D.H.;
RT "Genome organization of membrane-containing bacteriophage PRD1.";
RL Virology 183:658-676(1991).
CC -!- FUNCTION: THE MAJOR COAT PROTEIN P3 AND TWO ASSEMBLY FACTORS (P10
CC & P17) ARE NEEDED DURING THE ASSEMBLY OF THE VIRUS PARTICLE INSIDE
CC THE HOST CELL, WHEN THE CAPSID PROTEIN MULTIMERS ARE CAPABLE OF
CC ENCLOSING THE VIRAL MEMBRANE FROM THE HOST PLASMA MEMBRANE
CC CONTAINING THE VIRUS-ENCODED MEMBRANE-ASSOCIATED PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M69077; -; NOT_ANNOTATED_CDS.
CC PIR; E36777; WMBPTE.
CC Pfam; PF02178; AT_hook; 2.
SQ SEQUENCE 203 AA; 20688 MW; 064C2BB9C08151D9 CRC64;

Query Match 5.3%; Score 7; DB 1; Length 203;
Best Local Similarity 100.0%; Pred. No. 14; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY 83 AAPYIEQ 89
Db 190 AAPYIEQ 196
|||||||

RESULT 10
ISPD_BACHD
ID ISPD_BACHD STANDARD; PRT; 228 AA.
AC O9KGF8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (EC 2.7.7.60)
DE (4-diphosphocytidyl-2C-methyl-D-erythritol synthase) (MEP
DE cytidylyltransferase) (MCT).
GN ISPD OR BH0107.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C-125 / JCM 9153;
MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -!- FUNCTION: CATALYZES THE FORMATION OF 4-DIPHOSPHOCYTIDYL-2C-
CC METHYL-D-ERYTHRITOL FROM CTP AND 2C-METHYL-D-ERYTHRITOL 4-
CC PHOSPHATE (BY SIMILARITY).
```

```

CC -!- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate -
CC diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.
CC -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.
CC -!- SIMILARITY: BELONGS TO THE ISPD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP001507; BAB03826.1; -
CC PIR; C83663; C83663.
CC HSP; Q46893; I152.
CC HAMAP; MF_00108; -; 1.
CC InterPro; IPR001228; ISPD_synthase.
CC Pfam; PF01128; ISPD; 1.
CC PIRSF; PIRSF006765; DPCME_synth; 1.
CC TIGRFAMS; TIGR00453; ispd; 1.
CC PROSITE; PS01295; ISPD; 1.
CC Transferrase; Nucleotidyltransferase; Isoprene biosynthesis;
CC Complete proteome.
CC KW SEQUENCE 228 AA; 25270 MW; 1D94A1361DEDA080 CRC64;

Query Match 5.3%; Score 7; DB 1; Length 228;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 IELGGKP 61
Db 25 IELGGKP 31
|||||||

RESULT 11
YOGA_ECOLI
ID YOGA_ECOLI STANDARD; PRT; 235 AA.
AC Q46831;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yqga.
GN YOGA OR B2966.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U28377; AAA69134.1; -
CC EMBL; AE000379; AAC76003.1; -
CC PIR; E65082; E65082.
CC EcoGene; EG12987; yqga.
CC Pfam; PF04474; DUF554; 1.
```

KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 2 22 POTENTIAL.
FT TRANSSEM 34 54 POTENTIAL.
FT TRANSEM 56 76 POTENTIAL.
FT TRANSEM 102 122 POTENTIAL.
FT TRANSEM 147 167 POTENTIAL.
FT TRANSEM 178 198 POTENTIAL.
FT TRANSEM 210 230 POTENTIAL.
SQ SEQUENCE 235 AA; 24614 MW; CC89D2A93FED29EA CRC64;

Query Match 5.3%; Score 7; DB 1; Length 235;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 30 VLLGGVL 36
Db 12 VLLGGVL 18

RESULT 12
ISPD_RALSO STANDARD; PRT; 253 AA.
AC OBXW3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (MEP)
DE (4-diphosphocytidyl-2C-methyl-D-erythritol synthase) (MEP)
DE cytidyltransferase (MCTP).
GN ISPD OR RSC1643 OR RS04018.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=GM1000;
RX MEDLINE=21681879; PubMed=11823852;

RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottilier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
CC -1- FUNCTION: Catalyzes the formation of 4-diphosphocytidyl-2C-
methyl-D-erythritol from CTP and 2C-methyl-D-erythritol 4-
phosphate (By similarity).
CC -1- CATALYTIC ACTIVITY: CTP + 2-C-methyl-D-erythritol 4-phosphate =
diphosphate + 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol.
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; third step.
CC -1- SIMILARITY: BELONGS TO THE ISPD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL; AL64065; CAD15345.1; -;
CC HAMAP; MF_00108; -;
CC InterPro; IPR001228; ISPD_synthase.
CC Pfam; PF01128; ISPD; 1.
CC FIRST; PIRSF006765; DPCME_synth; 1.
CC TIGRFAMS; TIGR00453; ISPD; 1.
CC PROSITE; PS01295; ISPD; 1.
CC Transferase; Nucleotidyltransferase; Isoprene biosynthesis;
CC Complete proteome.
SQ SEQUENCE 253 AA; 26133 MW; 22CC2D6CFAE3B46E CRC64;

DR AL64065; CAD15345.1; -;
DR HAMAP; MF_00108; -;
DR InterPro; IPR001228; ISPD_synthase.
DR Pfam; PF01128; ISPD; 1.
DR FIRST; PIRSF006765; DPCME_synth; 1.
DR TIGRFAMS; TIGR00453; ISPD; 1.
DR PROSITE; PS01295; ISPD; 1.
DR Transferase; Nucleotidyltransferase; Isoprene biosynthesis;
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 253 AA; 26133 MW; 22CC2D6CFAE3B46E CRC64;

Query Match 5.3%; Score 7; DB 1; Length 253;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 102 LGLLQRA 108
Db 185 LGLLQRA 191

RESULT 13
SNAB_STRPR STANDARD; PRT; 276 AA.
AC PS4993;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Pristinamycin IIA synthase subunit B (PIIA synthase subunit B).
GN SNAB.
OS Streptomyces pristinaespiralis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=38300;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SP92;
RX MEDLINE=95394837; PubMed=7665509;
RA Blanc V., Lagneaux D., Didier P., Gil P., Lacroix P., Crouzet J.;
RT "Cloning and analysis of structural genes from Streptomyces
pristinaespiralis encoding enzymes involved in the conversion of
pristinamycin IIB to pristinamycin IIA (PIIA): PIIA synthase and
NADH:riboflavin 5'-phosphate oxidoreductase.";
RL J. Bacteriol. 177:5206-5214(1995).
RN [2]
RP SEQUENCE OF 1-22 AND 121-135.
RX MEDLINE=95394836; PubMed=7665508;

RA Thibaut D., Ratet N., Bisch D., Faucher D., Debussche L., Blanche F.;
RT "Purification of the two-enzyme system catalyzing the oxidation of
the D-proline residue of pristinamycin IIB during the last step of
pristinamycin IIA biosynthesis.";
RL J. Bacteriol. 177:5199-5205(1995).
CC -1- FUNCTION: CATALYZES THE OXIDATION OF THE PROLINE RESIDUE OF
PRISTINAMYCIN IIB (PIIB) TO PRISTINAMYCIN IIA (PIIA).
CC -1- COFACTOR: FMN.
CC -1- SUBUNIT: HETERODIMER OF TWO SUBUNITS, SNA AND SNAB.
CC -1- SIMILARITY: SOME SIMILARITY TO THE NTAA/SNAA/DSZA(SOXA) FAMILY OF
MONOOXYGENASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL; U21215; AAA83565.1; -;
CC KW Oxidoreductase; Monooxygenase; Flavoprotein; FMN.
FT INIT_MET 0
SQ SEQUENCE 276 AA; 28633 MW; EA8EA9E738D02E45 CRC64;

Query Match 5.3%; Score 7; DB 1; Length 276;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 35 VLAALAA 41
Db 239 VLAALAA 245

RESULT 14
AMPM_PYRFU STANDARD; PRT; 295 AA.
ID AMPM_PYRFU
AC P56218;

DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Methionine aminopeptidase (EC 3.4.11.18) (MAP) (Peptidase M).
 GN MAP OR PF0541.
 OS Pyrococcus furiosus.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC Pyrococcus.
 OX NCBI_TaxID=2261;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
 RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
 RT "The complete sequence of the Pyrococcus furiosus genome.";
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (1.75 ANGSTROMS).
 RX MEDLINE=98234705; PubMed=9573622;
 RA Tahirov T.H., Oki H., Tsukihara T., Ogasahara K., Yutani K.,
 RA Izu Y., Tsunasawa S., Kato I.;
 RT "High-resolution crystals of methionine aminopeptidase from
 Pyrococcus furiosus obtained by water-mediated transformation.";
 RL J. Struct. Biol. 121:68-72(1998).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (1.75 ANGSTROMS).
 RX MEDLINE=99030464; PubMed=9811545;
 RA Tahirov T.H., Oki H., Tsukihara T., Ogasahara K., Yutani K.,
 RA Izu Y., Tsunasawa S., Kato I.;
 RT "Crystal structure of methionine aminopeptidase from hyperthermophile,
 Pyrococcus furiosus.";
 RL J. Mol. Biol. 284:101-124(1998).
 CC -!- FUNCTION: REMOVES THE AMINO-TERMINAL METHIONINE FROM NASCENT
 PROTEINS (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: Release of N-terminal amino acids,
 preferentially methionine, from peptides and arylamides.
 CC -!- COFACTOR: COBALT; BINDS 2 IONS PER SUBUNIT.
 CC -!- SUBUNIT: Monomer.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24C.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AE010177; AAL80665.1; -.
 DR PIR: JC5671; JC5671.
 DR PDB: 1XGM; 25-FEB-98.
 DR PDB: 1XGN; 25-FEB-98.
 DR PDB: 1XGO; 25-FEB-98.
 DR PDB: 1XGS; 07-JUL-97.
 DR MEROPS: M24.002; -.
 DR InterPro: IPR002468; MAP 2.
 DR InterPro: IPR001714; Methamino_Ptase.
 DR InterPro: IPR000994; Peptidase_M24.
 DR Pfam: PF00557; Peptidase_M24; 1.
 DR PRINTS: PR00599; MAPEPTIDASE.
 DR TIGRfam: TIGR00501; met_pdae_II; 1.
 DR PROSITE: PS01202; MAP_2; 1.
 DR Hydrolase; Aminopeptidase; Cobalt; 3D-structure; Complete proteome.
 KW METAL 82 82
 FT METAL 93 93
 FT METAL 133 153
 FT METAL 187 187
 FT METAL 280 280
 FT METAL 3 23
 FT TURN 26 27
 FT STRAND 29 29
 FT HELIX 30 43
 FT TURN 44 45
 FT STRAND 47 48

FT STRAND 53 56
 FT TURN 57 58
 FT STRAND 59 61
 FT TURN 67 68
 FT STRAND 72 72
 FT TURN 75 76
 FT STRAND 78 87
 FT TURN 88 89
 FT STRAND 90 99
 FT TURN 100 101
 FT HELIX 106 121
 FT TURN 124 125
 FT STRAND 127 127
 FT TURN 128 128
 FT HELIX 129 140
 FT TURN 141 143
 FT STRAND 145 146
 FT TURN 148 149
 FT STRAND 152 154
 FT STRAND 156 156
 FT TURN 157 158
 FT STRAND 159 159
 FT STRAND 165 166
 FT TURN 172 173
 FT STRAND 177 177
 FT TURN 180 181
 FT STRAND 183 186
 FT STRAND 189 191
 FT STRAND 197 208
 FT HELIX 217 229
 FT TURN 230 232
 FT STRAND 235 236
 FT HELIX 238 240
 FT TURN 241 243
 FT HELIX 246 259
 FT TURN 260 260
 FT STRAND 262 270
 FT TURN 271 272
 FT STRAND 276 278
 FT STRAND 280 285
 FT STRAND 290 292
 FT TURN 293 295
 SQ SEQUENCE 295 AA; 32842 MW; 9739BC55F812E65B CRC64;
 Query Match 5.3%; Score 7; DB 1; Length 295;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 56 ELGGKPA 62
 Db 43 ELGGKPA 49
 RESULT 15
 AMPM_PVRHO STANDARD; PRT; 295 AA.
 AC O58362;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Methionine aminopeptidase (EC 3.4.11.18) (MAP) (Peptidase M).
 GN MAP OR PH0628
 OS Pyrococcus horikoshii.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC Pyrococcus.
 OX NCBI_TaxID=53953;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OT3;
 RX MEDLINE=98344137; PubMed=9679194;
 RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
 RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
 RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,

RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
CC -1- FUNCTION: REMOVES THE AMINO-TERMINAL METHIONINE FROM NASCENT
CC PROTEINS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Release of N-terminal amino acids,
CC preferentially methionine, from peptides and arylamides.
CC -1- COFACTOR: COBALT; BINDS 2 IONS PER SUBUNIT (By similarity).
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP000003; BAA29717.1; -;
DR PIR: C71107; C71107.
DR HSP; P56218; IXGS.
DR MEROPS; M24.002; -;
DR InterPro; IPR002468; MAP_2.
DR InterPro; IPR001714; Methamino_Ptase.
DR InterPro; IPR000994; Peptidase_M24.
DR Pfam; PF00557; Peptidase_M24; 1.
DR PRINTS; PR00599; MAPEPTIDASE.
DR TIGRFAMS; TIGR00501; met_pdase_II; 1.
DR PROSITE; PS01202; MAP_2; 1.
DR Hydrolase; Amino-peptidase; Cobalt; Complete proteome.
KW METAL 82 82 COBALT 2 (By similarity).
FT METAL 93 93 COBALT 1 AND 2 (By similarity).
FT METAL 153 153 COBALT 1 (By similarity).
FT METAL 187 187 COBALT 1 (By similarity).
FT METAL 280 280 COBALT 1 AND 2 (By similarity).
SQ SEQUENCE 295 AA; 32795 MW; D228F4377CEB2RAC CRC64;

Query Match 5.3%; Score 7; DB 1; Length 295;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 56 ELGGKPA 62
| | | | | | |
Db 43 ELGGKPA 49

Search completed: August 29, 2003, 11:24:22
Job time : 14 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 11:19:39 ; Search time 29 Seconds
(without alignments)
1183.483 Million cell updates/sec

Title: US-09-638-693-36
Perfect score: 133
Sequence: 1 QNEICLTHPTIKYIMACMSA.....VIEPIVTNMQKLEAFWHKH 133

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*
15: sp-rvirus:*
16: sp-bacteriaph:*
17: sp-archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	98	73.7	209	12 Q81594	Q81594 hepatitis c
2	52	39.1	133	12 Q81595	Q81595 hepatitis c
3	52	39.1	3021	12 Q68870	Q68870 hepatitis c
4	48	36.1	3021	12 Q92933	Q92933 hepatitis c
5	43	32.3	3021	12 Q81258	Q81258 hepatitis c
6	35	26.3	70	12 Q39899	Q39899 hepatitis c
7	35	26.3	88	12 Q39898	Q39898 hepatitis c
8	35	26.3	90	12 Q39904	Q39904 hepatitis c
9	35	26.3	90	12 Q39906	Q39906 hepatitis c
10	35	26.3	90	12 Q39908	Q39908 hepatitis c
11	35	26.3	90	12 Q39910	Q39910 hepatitis c
12	35	26.3	138	12 Q68239	Q68239 hepatitis c
13	32	24.1	83	12 Q39900	Q39900 hepatitis c
14	31	23.3	84	12 Q39896	Q39896 hepatitis c
15	31	23.3	84	12 Q39917	Q39917 hepatitis c
16	30	22.6	3023	12 Q81487	Q81487 hepatitis c

17	29	21.8	89	12 Q39894	Q39894 hepatitis c
18	29	21.8	89	12 Q39893	Q39893 hepatitis c
19	29	21.8	90	12 Q39905	Q39905 hepatitis c
20	28	21.1	84	12 Q39903	Q39903 hepatitis c
21	28	21.1	90	12 Q39916	Q39916 hepatitis c
22	28	21.1	90	12 Q39902	Q39902 hepatitis c
23	28	21.1	90	12 Q39901	Q39901 hepatitis c
24	28	21.1	90	12 Q39912	Q39912 hepatitis c
25	28	21.1	90	12 Q39913	Q39913 hepatitis c
26	28	21.1	138	12 Q68241	Q68241 hepatitis c
27	28	21.1	138	12 Q68233	Q68233 hepatitis c
28	24	18.0	3021	12 Q81495	Q81495 hepatitis c
29	22	16.5	90	12 Q39909	Q39909 hepatitis c
30	21	15.8	87	12 Q39907	Q39907 hepatitis c
31	20	15.0	193	12 Q56637	Q56637 hepatitis c
32	19	14.3	138	12 Q68209	Q68209 hepatitis c
33	19	14.3	138	12 Q68235	Q68235 hepatitis c
34	19	14.3	138	12 Q68218	Q68218 hepatitis c
35	19	14.3	138	12 Q68244	Q68244 hepatitis c
36	19	14.3	138	12 Q68242	Q68242 hepatitis c
37	19	14.3	138	12 Q68240	Q68240 hepatitis c
38	19	14.3	138	12 Q68221	Q68221 hepatitis c
39	19	14.3	138	12 Q68215	Q68215 hepatitis c
40	19	14.3	138	12 Q68232	Q68232 hepatitis c
41	19	14.3	138	12 Q68208	Q68208 hepatitis c
42	19	14.3	138	12 Q68231	Q68231 hepatitis c
43	19	14.3	138	12 Q68211	Q68211 hepatitis c
44	19	14.3	138	12 Q68237	Q68237 hepatitis c
45	19	14.3	138	12 Q68217	Q68217 hepatitis c

ALIGNMENTS

RESULT 1

Q81594 ID Q81594 PRELIMINARY; PRT; 209 AA.
AC Q81594;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Nonstructural protein 4 (Fragment).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93249436; PubMed=7683463;
RA Stuyver L., Arnhem W.V., Wyseur A., Deleys R., Maertens G.;
RT Analysis of the putative E1 envelope and NS4a epitope region of HCV type 3.*;
RT Biochem. Biophys. Res. Commun. 192:635-641(1993).
DR EMBL; D14600; BAA03449.1;
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001;
DR MEROPS; U39.001;
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1
FT NON_TER 209
SQ SEQUENCE 209 AA; 23408 MW; 76648D9BB1D3CD12 CRC64;

Query Match 73.7%; Score 98; DB 12; Length 209;

Best Local Similarity 100.0%; Pred. No. 5e-86;

Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QNEICLTHPTIKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSGCVVIVGHIELGK 60
|||||

DB 77 QNEICLTHPTIKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSGCVVIVGHIELGK 136
|||||

QY 61 PAIVDPKEVLYQOYDEMECSQAAPYIEQAQVIAHQFK 98

|||||
137 PAIYPDREVLYQYDEMECSQAAPYIEQAQVIAHQK 174

RESULT 2

Q81595 AC Q81595 PRELIMINARY; PRT; 133 AA.
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Nonstructural protein 4 (Fragment).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93249436; PubMed=7683463;
RA Stuyver L., Arnhem W.V., Wyseur A., Deleys R., Maertens G.;
RT "Analysis of the putative E1 envelope and NS4a epitope region of HCV
type 3";
RL Biochem. Biophys. Res. Commun. 192:635-641(1993).
DR EMBL; D14602; BAA03451.1; -;
DR HSSP; P27958; 1HEI.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1
FT NON_TER 133 133
SQ SEQUENCE 133 AA; 14781 MW; 4BFF2128FD301691 CRC64;

Query Match 39.1%; Score 52; DB 12; Length 133;
Best Local Similarity 100.0%; Pred. No. 5e-43;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 TKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSGVGVVIVGHIELGKPA 62
|||||
Db 11 TKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSGVGVVIVGHIELGKPA 62

RESULT 3

Q68870 ID Q68870 PRELIMINARY; PRT; 3021 AA.
AC Q68870;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genes for core, envelope and NS1 proteins (Genome
polyprotein).
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Seelig R., Weber P., Seeling H.P., Ledger N., Bottger C., Renz M.;
RT "Hepatitis C virus type V genome isolated from a patient in Germany";
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; X76918; CAA54244.1; -;
DR HSSP; P27958; 1A1V.
DR MEROPS; S29.001; -;
DR MEROPS; U39.001; -;
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.

DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RDRP.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; Viral_RDRP; 1.
DR Pfam; PF0186062; HCV_NS1; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS05057; RDRP_POSITIVE; 1.
DR PROSITE; PS05021; RDRP_VIRAL; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
FT CHAIN 1 191
FT CHAIN 192 383
FT CHAIN 384 735
FT CHAIN 736 1012
FT CHAIN 1013 1663
FT CHAIN 1664 1717
FT CHAIN 1718 1978
FT CHAIN 1979 2430
FT CHAIN 2431 3021
SQ SEQUENCE 3021 AA; 329092 MW; BF2B499AA55A58CF CRC64;

Query Match 39.1%; Score 52; DB 12; Length 3021;
Best Local Similarity 100.0%; Pred. No. 7.5e-42;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 TKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSGVGVVIVGHIELGKPA 62
|||||
Db 1648 TKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSGVGVVIVGHIELGKPA 1699

RESULT 4

O92933 ID O92933 PRELIMINARY; PRT; 3021 AA.
AC O92933;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-type 3a;
RA Shukla D.D., Chaturvedi S., Cao J.Y., Hoynes P.A.;
RT "Complete Nucleotide Sequence of the genome of Hepatitis C virus type
3a (Cb)";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF046866; AAC03058.1; -;
DR HSSP; P27958; 1HEI.
DR MEROPS; S29.001; -;
DR MEROPS; U39.001; -;
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.

```
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR00745; HCV_NS4a.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RDRP.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RDRP.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00435; PEROXIDASE_1; 1.
DR PROSITE: PS0507; RDRP_POSITIVE; 1.
DR PROSITE: PS0521; RDRP_VIRAL; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3021 AA; 328903 MW; D7B6133B33030303CD CRC64;

Query Match 36.1%; Score 48; DB 12; Length 3021;
Best Local Similarity 100.0%; Pred. No. 6.3e-38;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 MACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGKPA 62
    |||||
Db 1652 MACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGKPA 1699

RESULT 5
Q81258 PRELIMINARY; PRT; 3021 AA.
AC Q81258;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NZL1;
RA Okamoto H.;
RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NZL1;
RA Sakamoto M.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: D17763; BAA04609.1; -.
DR HSP: P27958; IHEI.
DR MEROPS: S29.001; -.
DR MEROPS: U39.001; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002531; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
```

```
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR00745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RDRP.
DR InterPro: IPR002016; Peroxidase.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01560; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS00435; PEROXIDASE_1; 1.
DR PROSITE: PS0507; RDRP_POSITIVE; 1.
DR PROSITE: PS0521; RDRP_VIRAL; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
FT CHAIN 1 191
FT CHAIN 192 383
FT CHAIN 384 735
FT CHAIN 736 1012
FT CHAIN 1013 1663
FT CHAIN 1664 1717
FT CHAIN 1718 1978
FT CHAIN 1979 2430
FT CHAIN 2431 3021
SQ SEQUENCE 3021 AA; 329574 MW; 38712CCBC0C19562 CRC64;

Query Match 32.3%; Score 43; DB 12; Length 3021;
Best Local Similarity 100.0%; Pred. No. 5.1e-33;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 MACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIEL 57
    |||||
Db 1652 MACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIEL 1694

RESULT 6
Q39899 PRELIMINARY; PRT; 70 AA.
AC Q39899;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Non-structural protein 4a/b (Fragment).
GN NS4A/B.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GD.1.3;
RX MEDLINE=98032593; PubMed=9365889;
RA Prescott L.E., Berger A., Pawlotsky J.M., Conjeevaram P., Pike I.,
RA Simmonds P.;
RT "Sequence analysis of hepatitis C virus variants producing discrepant
RT results with two different genotyping assays.";
RL J. Med. Virol. 53:237-244(1997).
DR EMBL: AF007504; AAB62955.2; -.
DR InterPro: IPR000745; HCV_NS4a.
DR Pfam: PF01006; HCV_NS4a; 1.
DR NON_TER 1
FT NON_TER 70 70
SQ SEQUENCE 70 AA; 7743 MW; 818D296E0E488DB1 CRC64;
```

Query Match 26.3%; Score 35; DB 12; Length 70;
Best Local Similarity 100.0%; Pred. No. 1.4e-26;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 VPDKEVLYQQYDEMECSQAAPYIEQAQVIAHQFK 98
|||||
DB 22 VPDKEVLYQQYDEMECSQAAPYIEQAQVIAHQFK 56
|||||

RESULT 7

O39898 ID O39898 PRELIMINARY; PRT; 88 AA.

AC O39898; 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JUN-2002 (TReMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Non-structural protein 4a/b (Fragment).
GN NS4A/B.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GD_1.1;
RX MEDLINE=98032593; PubMed=9365889;
RA Prescott L.E., Berger A., Pawlotsky J.M., Conjeevaram P., Pike I.,
RA Simmonds P.;
RT "Sequence analysis of hepatitis C virus variants producing discrepant
RT results with two different genotyping assays.";
RL J. Med. Virol. 53:237-244(1997).
DR EMBL; AF007503; AAB62954.2; -.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1
FT NON_TER 88
SQ SEQUENCE 88 AA; 9838 MW; 91FDFBC4EDD171FF CRC64;

Query Match 26.3%; Score 35; DB 12; Length 88;
Best Local Similarity 100.0%; Pred. No. 1.7e-26;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 VPDKEVLYQQYDEMECSQAAPYIEQAQVIAHQFK 98
|||||
DB 23 VPDKEVLYQQYDEMECSQAAPYIEQAQVIAHQFK 57
|||||

RESULT 8

O39904 ID O39904 PRELIMINARY; PRT; 90 AA.

AC O39904; 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Non-structural protein 4a/b (Fragment).
GN NS4A/B.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Prescott L.E., Berger A., Pawlotsky J.M., Conjeevaram P., Pike I.,
RA Simmonds P.;
RT "Sequence analysis of Hepatitis C virus variants producing discrepant
RT results with two different genotyping assays.";
RL J. Med. Virol. 0:0-0(1997).
DR EMBL; AF007509; AAB62960.1; -.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1
FT NON_TER 90
SQ SEQUENCE 90 AA; 9982 MW; D4E5CD868E87769C CRC64;

SQ SEQUENCE .90 AA; 10000 MW; 15465423A211108B CRC64;

Query Match 26.3%; Score 35; DB 12; Length 90;
Best Local Similarity 100.0%; Pred. No. 1.7e-26;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 VPDKEVLYQQYDEMECSQAAPYIEQAQVIAHQFK 98
|||||
DB 25 VPDKEVLYQQYDEMECSQAAPYIEQAQVIAHQFK 59
|||||

RESULT 9

O39906 ID O39906 PRELIMINARY; PRT; 90 AA.

AC O39906; 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Non-structural protein 4a/b (Fragment).
GN NS4A/B.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Prescott L.E., Berger A., Pawlotsky J.M., Conjeevaram P., Pike I.,
RA Simmonds P.;
RT "Sequence analysis of Hepatitis C virus variants producing discrepant
RT results with two different genotyping assays.";
RL J. Med. Virol. 0:0-0(1997).
DR EMBL; AF007511; AAB62962.1; -.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1
FT NON_TER 90
SQ SEQUENCE 90 AA; 10031 MW; DC7CDBE6E881109C CRC64;

Query Match 26.3%; Score 35; DB 12; Length 90;
Best Local Similarity 100.0%; Pred. No. 1.7e-26;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 VPDKEVLYQQYDEMECSQAAPYIEQAQVIAHQFK 98
|||||
DB 25 VPDKEVLYQQYDEMECSQAAPYIEQAQVIAHQFK 59
|||||

RESULT 10

O39908 ID O39908 PRELIMINARY; PRT; 90 AA.

AC O39908; 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Non-structural protein 4a/b (Fragment).
GN NS4A/B.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RA Prescott L.E., Berger A., Pawlotsky J.M., Conjeevaram P., Pike I.,
RA Simmonds P.;
RT "Sequence analysis of Hepatitis C virus variants producing discrepant
RT results with two different genotyping assays.";
RL J. Med. Virol. 0:0-0(1997).
DR EMBL; AF007513; AAB62964.1; -.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1
FT NON_TER 90
SQ SEQUENCE 90 AA; 9982 MW; D4E5CD868E87769C CRC64;

Query Match 26.3%; Score 35; DB 12; Length 90;
 Best Local Similarity 100.0%; Pred. No. 1.7e-26;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 VPDKEVLYQYDEMEECQAAPYIEQAQVIAHQFK 98
 |||||
 DB 25 VPDKEVLYQYDEMEECQAAPYIEQAQVIAHQFK 59

RESULT 11

O39910
 ID O39910 PRELIMINARY; PRT; 90 AA.
 AC O39910;
 DT 01-JAN-1998 (TReMBLrel. 05, Created)
 DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE Non-structural protein 4a/b (Fragment).
 GN NS4A/B.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=111103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Prescott L.E., Berger A., Pawlotsky J.-M., Conjeevaram P., Pike I.,
 RA Simmonds P.;
 RT "Sequence analysis of Hepatitis C virus variants producing discrepant
 RT results with two different genotyping assays.";
 RL J. Med. Virol. 0:0-0(1997).
 DR EMBL: AF007515; AAB62966.1; -.
 DR InterPro: IPR000745; HCV_NS4a.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR NON_TER 1
 FT NON_TER 90
 SQ SEQUENCE 90 AA; 9982 MW; D4E5CD868E87769C CRC64;

Query Match 26.3%; Score 35; DB 12; Length 90;
 Best Local Similarity 100.0%; Pred. No. 1.7e-26;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 VPDKEVLYQYDEMEECQAAPYIEQAQVIAHQFK 98
 |||||
 DB 25 VPDKEVLYQYDEMEECQAAPYIEQAQVIAHQFK 59

RESULT 12

Q68239
 ID Q68239 PRELIMINARY; PRT; 138 AA.
 AC Q68239;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE Nonstructural protein (Fragment).
 GN NS4.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=111103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3a;
 RX MEDLINE=95146953; PubMed=7844535;
 RA Greene W.K., Cheong M.K., Ng V., Yap K.W.;
 RT "Prevalence of hepatitis C virus sequence variants in South-East
 RT Asia.";
 RL J. Gen. Virol. 76:211-215(1995).
 DR EMBL: U14275; AAC53964.1; -.
 DR HSP: P27958; LHET.
 DR InterPro: IPR000745; HCV_NS4a.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR NON_TER 1
 FT NON_TER 138 138

SQ SEQUENCE 138 AA; 15311 MW; B45AC0F8917DBAAC CRC64;

Query Match 26.3%; Score 35; DB 12; Length 138;
 Best Local Similarity 100.0%; Pred. No. 2.5e-26;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 VPDKEVLYQYDEMEECQAAPYIEQAQVIAHQFK 98
 |||||
 DB 59 VPDKEVLYQYDEMEECQAAPYIEQAQVIAHQFK 103

RESULT 13

O39900
 ID O39900 PRELIMINARY; PRT; 83 AA.
 AC O39900;
 DT 01-JAN-1998 (TReMBLrel. 05, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
 DE Non-structural protein 4a/b (Fragment).
 GN NS4A/B.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=111103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Prescott L.E., Berger A., Pawlotsky J.-M., Conjeevaram P., Pike I.,
 RA Simmonds P.;
 RT "Sequence analysis of Hepatitis C virus variants producing discrepant
 RT results with two different genotyping assays.";
 RL J. Med. Virol. 0:0-0(1997).
 DR EMBL: AF007505; AAB62956.1; -.
 DR InterPro: IPR000745; HCV_NS4a.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR NON_TER 1
 FT NON_TER 83
 SQ SEQUENCE 83 AA; 9089 MW; 0FAA36D622B24BDE CRC64;

Query Match 24.1%; Score 32; DB 12; Length 83;
 Best Local Similarity 100.0%; Pred. No. 1.4e-23;
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 KEVLYQYDEMEECQAAPYIEQAQVIAHQFK 98
 |||||
 DB 28 KEVLYQYDEMEECQAAPYIEQAQVIAHQFK 59

RESULT 14

O39896
 ID O39896 PRELIMINARY; PRT; 84 AA.
 AC O39896;
 DT 01-JAN-1998 (TReMBLrel. 05, Created)
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
 DE Non-structural protein 4a/b (Fragment).
 GN NS4A/B.
 OS Hepatitis C virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OC NCBI_TaxID=111103;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GD_7;
 RX MEDLINE=98032593; PubMed=9365889;
 RA Prescott L.E., Berger A., Pawlotsky J.M., Conjeevaram P., Pike I.,
 RA Simmonds P.;
 RT "Sequence analysis of hepatitis C virus variants producing discrepant
 RT results with two different genotyping assays.";
 RL J. Med. Virol. 53:237-244(1997).
 DR EMBL: AF007501; AAB62952.2; -.
 DR InterPro: IPR000745; HCV_NS4a.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR NON_TER 1
 FT NON_TER 1

```
FT NON_TER      84
SQ SEQUENCE      84 AA; 9343 MW; 135FFAB56ADD4D16 CRC64;

Query Match      23.3%; Score 31; DB 12; Length 84;
Best Local Similarity 100.0%; Pred. No. 1.4e-22;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 68 EVLYQQYDEMEECSSQAAPYIEQAQVIAHQFK 98
Db 28 EVLYQQYDEMEECSSQAAPYIEQAQVIAHQFK 58

RESULT 15
O39917 PRELIMINARY; PRT; 84 AA.
AC O39917;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Non-structural protein 4a/b (fragment).
GN NS4A/B.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FD_4;
RX MEDLINE=98032593; PubMed=9365889;
RA Prescott L.E., Berger A., Pawlotsky J.M., Conjeevaram P., Pike I.,
RA Simmonds P.;
RT "Sequence analysis of hepatitis C virus variants producing discrepant
RT results with two different genotyping assays.";
RL J. Med. Virol. 53:237-244(1997).
DR EMBL: AF007522; AAB62973.2; -.
DR InterPro: IPR00745; HCV_NS4a.
DR Pfam: PF01006; HCV_NS4a; 1.
FT NON_TER      1
FT NON_TER      84
SQ SEQUENCE      84 AA; 9360 MW; 88E9F6DBE6E66DA7 CRC64;

Query Match      23.3%; Score 31; DB 12; Length 84;
Best Local Similarity 100.0%; Pred. No. 1.4e-22;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 68 EVLYQQYDEMEECSSQAAPYIEQAQVIAHQFK 98
Db 24 EVLYQQYDEMEECSSQAAPYIEQAQVIAHQFK 54
```

Search completed: August 29, 2003, 11:25:01
Job time : 31 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 29, 2003, 12:52:20 ; Search time 1822 Seconds
(without alignments)
1774.147 Million cell updates/sec

Title: US-09-638-693-36
Perfect score: 133
Sequence: 1 QNEICLTHPTKYIMACMSA.....VIEPIVTNNQKLEAFWHKH 133

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Word size: 1

Total number of hits satisfying chosen parameters: 45562604

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL-frame+p2n.model -DEV-rlx
-Q/cgr2_1/USPTO.spool/US09638693/runat_25082003_173340_22273/app.query.fasta_1.327
-DB-EST -QFMT-fastap -SUFFIX=oli.rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS-human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09638693 @CN_1_1_2810 @runat_25082003_173340_22273 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : EST.*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_est1.*
10: gb_est2.*
11: gb_hc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	9	6.8	399	10	BE653925 UI-M-AIO-
C 2	9	6.8	427	13	BY275359 BY275359
C 3	9	6.8	448	10	BE647694 UI-M-BH1-
C 4	9	6.8	469	14	CA801050 sau23h07-
C 5	9	6.8	552	13	BQ555344 H4033E09-
C 6	9	6.8	631	14	BY727796 UI-M-FRO-
C 7	9	6.8	737	13	BU707252 UI-M-FRO-
C 8	9	6.8	743	12	BM951073 UI-M-EGO-
C 9	9	6.8	758	13	BU707783 UI-M-FRO-
C 10	9	6.8	799	12	BI968400 GM830005A
C 11	9	6.8	889	13	BX425691 BX425691
C 12	9	6.8	904	13	BQ715452 AGENCOURT
C 13	9	6.8	1016	13	BQ900568 AGENCOURT
C 14	9	6.8	1101	29	CNS05658 Tetraodon
C 15	9	6.8	1142	14	CD496483 CDA22-E10
C 16	8	6.0	71	28	AQ846889 LMAJFV1_1
C 17	8	6.0	164	9	AV638488 AV638488
C 18	8	6.0	173	9	AV387996 AV387996
C 19	8	6.0	186	9	AV637755 AV637755
C 20	8	6.0	207	9	AA652442 ns60h05.s
C 21	8	6.0	217	28	BH299720 CH230-71P
C 22	8	6.0	218	9	AV389274 AV389274
C 23	8	6.0	229	9	AV389781 AV389781
C 24	8	6.0	241	9	AV640464 AV640464
C 25	8	6.0	242	9	AV388321 AV388321
C 26	8	6.0	249	9	AV389230 AV389230
C 27	8	6.0	250	9	AV388577 AV388577
C 28	8	6.0	253	9	AV630624 AV630624
C 29	8	6.0	260	9	AV389975 AV389975
C 30	8	6.0	276	13	BY275822 BY275822
C 31	8	6.0	277	9	AW210711 um64609.x
C 32	8	6.0	279	9	AV252548 AV252548
C 33	8	6.0	280	14	CB964946 NL14_B06
C 34	8	6.0	284	9	AV388546 AV388546
C 35	8	6.0	287	28	BH645681 BOMHT73TR
C 36	8	6.0	292	9	AV627125 AV627125
C 37	8	6.0	292	9	AV627904 AV627904
C 38	8	6.0	296	9	AV631317 AV631317
C 39	8	6.0	303	12	BI641192 SD24407.5
C 40	8	6.0	303	13	BQ809739 1030012H0
C 41	8	6.0	309	9	AV628578 AV628578
C 42	8	6.0	310	9	AV388479 AV388479
C 43	8	6.0	312	28	B62850 CIT978SK-22
C 44	8	6.0	318	9	AV645205 AV645205
C 45	8	6.0	319	9	AV388996 AV388996

ALIGNMENTS

RESULT 1
BE653925/c
LOCUS UI-M-AIO-aal-a-12-0-UI.r1 NIH_BMAP_MBS Mus musculus cDNA clone
DEFINITION UI-M-AIO-aal-a-12-0-UI 5', mRNA sequence.
ACCESSION BE653925
VERSION BE653925.1 GI:9979838
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 399)

AUTHORS Ronaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Chin, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: mestr@mail.nih.gov
 cDNA Library Preparation: M.B. Soares Lab Clone Distribution:
 Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
 should be noted that Bento Soares is generating a small number of
 additional specialized non-redundant arrays of BMAP cDNAs whose
 availability will be considered under appropriate and limited
 collaborative arrangements. The following repetitive elements were
 found in this cDNA sequence: 45-106, >(CAG)n#simple-repeat 262-305,
 >GC-rich#low-complexity
 Seq primer: M13 Reverse.
FEATURES Location/Qualifiers
 source
 1..399
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-A10-aal-a-12-0-UI"
 /dev_stage="27-32 days"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NIH_BMAP_MBS"
 /note="vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; Site.1: Not I; Site.2: Eco RI; The
 NIH_BMAP_MBS library is a non-normalized library
 constructed from mouse brain stems. The tag is a string
 of 5 nucleotides present between the Not I site and the
 oligo-dr track. The library was constructed as described
 by Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
 1996. Tissue provided by Ms. Annie Novakovich,
 Zivic-Miller Laboratories."
 BASE COUNT 60 a 152 c 134 g 53 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 179 Length: 399
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 6.77% Indels: 0
 DB: 10 Gaps: 0
 US-09-638-693-36 (1-133) x BE653925 (1-399)
 Qy 104 LeuLeuGlnArgAlaThrGlnGln 112
 Db 259 CTCCTGCAGCGAGCTACCCAGCAGCAG 233
 RESULT 2
 BY275359/c 427 bp mRNA linear EST 11-DEC-2002
 LOCUS
 DEFINITION BY275359 RIKEN full-length enriched, visual cortex Mus musculus
 cDNA clone K430312004 5', mRNA sequence.
 ACCESSION BY275359
 VERSION BY275359.1 GI:26465696
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 427)
 Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
 Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.

, Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
 Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
 Quackenbush,J., Schreli,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
 Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,
 L.E., Cousins,S., Dalia,E., Dragani,T.A., Fletcher,C.F., Forrest,
 A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.,
 Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
 Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
 King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,
 P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,
 H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Perle,G.,
 Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
 Ramachandran,S., Ravasi,T., Reed,J.C., Reed,J.U., Reid,J., Ring,
 B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,
 M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,
 R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
 Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yangisawa,
 M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
 Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,
 M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
 Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,
 Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,
 K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,
 E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 22354683
 12466851
COMMENT Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gs.riken.go.jp/
 URL:http://genome.gsc.riken.go.jp/
 Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,
 T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A.,
 Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K.,
 Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,
 M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
 Submission
 Computational Analysis of Full-Length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissues were provided by Michela Fagiolini and Takao K. Hensch (
 Laboratory for Neuronal Circuit Development Brain Science Institute
 RIKEN 2-1 Hirosewa,Wako-Shi,Saitama 351-0198 Japan) whose
 assistance we gratefully acknowledge. Please visit our web site
 (http://genome.gsc.riken.go.jp) for further details.
 Location/Qualifiers
 1..427
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="K430312004"
 /tissue_type="visual cortex"

FEATURES
 source

```

/clone_lib="RIKEN full-length enriched, visual cortex"
BASE COUNT      65 a 162 c 139 g 60 t 1 others
ORIGIN

Alignment Scores:
Pred. No.:      192      Length:      427
Score:          9.00      Matches:      9
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:     6.77%      Indels:      0
DB:              13      Gaps:        0

US-09-638-693-36 (1-133) x BY275359 (1-427)

QY      104 LeuLeuGlnArgAlaThrGlnGln 112
Db      274 CTCCTGCAGCAGCTACCCAGCAGCAG 248

RESULT 3
LOCUS      BE647694/c
DEFINITION      UI-M-BH1-ano-a-08-0-UI.r1 NIH_BMAP_M_S2 Mus musculus cDNA clone
ACCESSION      BE647694
VERSION        BE647694.1 GI:9973507
KEYWORDS      EST.
SOURCE        Mus musculus (house mouse)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 448)
AUTHORS      Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE      Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL      Genome Res. 6 (9), 791-806 (1996)
MEDLINE      97044477
PUBMED      889548
COMMENT      Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20832-3643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mestr@mail.nih.gov
cDNA Library Preparation: M.B. Soares Lab Clone distribution:
Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
should be noted that Bento Soares is generating a small number of
additional specialized non-redundant arrays of BMAP cDNAs whose
availability will be considered under appropriate and limited
collaborative arrangements. The following repetitive elements were
found in this cDNA sequence: 103-164, >(CAG)n#Simple_repeat 320-363
>GC-rich/low complexity
Seq primer: M13 Reverse.
Location/Qualifiers
1..448
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH1-ano-a-08-0-UI"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NIH_BMAP_M_S2"
/note="Vector: pTT3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not 1; Site_2: Eco RI; The
NIH_BMAP_M_S2 library is a subtracted library derived from
NIH_BMAP_M_S1, which in turn is a subtracted library
derived from a mixture of normalized libraries from ten
regions of the mouse brain (cerebellum, brain stems,
olfactory bulbs, hypothalamus, cortex, amygdala, basal
ganglia, pineal gland, striatum, hippocampus). The driver
used for subtraction consisted of a pool of 5,000 clones
from the NIH_BMAP_M_S1 library and a pool of 2,000 clones

```

```

obtained from non-normalized and normalized mouse brain
spinal cord libraries."
BASE COUNT      69 a 171 c 145 g 62 t 1 others
ORIGIN

Alignment Scores:
Pred. No.:      202      Length:      448
Score:          9.00      Matches:      9
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:     6.77%      Indels:      0
DB:              10      Gaps:        0

US-09-638-693-36 (1-133) x BE647694 (1-448)

QY      104 LeuLeuGlnArgAlaThrGlnGln 112
Db      317 CTCCTGCAGCAGCTACCCAGCAGCAG 291

RESULT 4
LOCUS      CA801050
DEFINITION      sau23h07.y1 Gm-cl062 Glycine max cDNA clone SOYBEAN CLONE ID:
Gm-cl062-9205 5' similar to TR:Q41102 Q41102 PHASEOLIN G-BOX
BINDING PROTEIN PG2 ;, mRNA sequence.
ACCESSION      CA801050
VERSION        CA801050.1 GI:26058136
KEYWORDS      EST.
SOURCE        Glycine max (soybean)
ORGANISM      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE      1 (bases 1 to 469)
AUTHORS      Shoemaker,R., Keim,P., Vodkin,L., Erpellding,J., Coryell,V., Khanna
,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Resgen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact: ccu@resgen.com web site:
www.resgen.com
Seq primer: -40RP from Gibco
High quality sequence stop: 377.
Location/Qualifiers
1..469
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-cl062-9205"
/tissue_type="stem tissue of greenhouse grown plants"
/dev_stage="1 month old"
/lab_host="DH10B"
/clone_lib="Gm-cl062"
/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
XhoI; The cDNA library was constructed from mRNA isolated
from stem tissue of 1 month old greenhouse grown plants
for the cultivar Raiden. Complementary DNA was
synthesized from mRNA using a primer consisting of a
poly(dT) sequence with a XhoI restriction site. EcoRI
adapters were ligated to the blunt-ended cDNA fragments

```

followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy Shoemaker."

BASE COUNT
ORIGIN

87 a 142 c 134 g 106 t

Alignment Scores:

Pred. No.: 212 Length: 469
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.77% Indels: 0
DB: 14 Gaps: 0

US-09-638-693-36 (1-133) x CA801050 (1-469)

QY 25 ThrThrSerThrTrpValLeuLeuGly 33
DB 309 ACAACTCGACATGGTCTCTTGGC 335

RESULT 5

BQ555344/c

LOCUS

DEFINITION BQ555344 552 bp mRNA linear EST 20-JUN-2002
H4033E09 5', mRNA sequence.

ACCESSION

VERSION BQ555344

KEYWORDS

SOURCE EST. GI:21456232

ORGANISM

Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 552)
VanBuren, V., Plao, Y., Dudekula, D.B., Qian, Y., Carter, M.G., Martin
P.R., Stagg, C.A., Bassey, D., Alba, K., Hamatani, T., Kargul, G.J.,
Luo, A.G., Kelso, J., Hide, W. and Ko, M.S.H.
Assembly, verification, and initial annotation of NIA 7.4K mouse
cDNA clone set

TITLE

- JOURNAL Genome Res. 12 (12); 1999-2003 (2002)
MEDLINE 22354164
PUBMED 12466305

COMMENT

Other ESTs: H4033E09-3

Contact: Yong Qian

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA

Email: cdna@igsun.grc.nia.nih.gov

This clone set has been freely distributed to the community. Please
visit http://igsun.grc.nia.nih.gov/cDNA/NIA_7.4k.html for details.

Plate: H4033 row: E column: 09

Seq primer: -21M13 Reverse

High quality sequence stop: 552

POLYA-No.

FEATURES

source

Location/Qualifiers
1..552
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="niaEST:H4033E09-5"
/db_xref="taxon:10090"
/clone="H4033E09"
/sex="mixed"
/dev_stage="mixed"
/lab_host="DH10B"
/clone_lib="NIA Mouse 7.4K cDNA Clone Set"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
clone is among a rearranged set of 7,407 clones from more
than 20 cDNA libraries."

BASE COUNT

ORIGIN

73 a 206 c 182 g 91 t

Alignment Scores:

Pred. No.: 251 Length: 552
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.77% Indels: 0
DB: 13 Gaps: 0

US-09-638-693-36 (1-133) x BQ555344 (1-552)

QY 104 LeuLeuGlnArgAlaThrGlnGlnGln 112

DB 219 CTCCTGCAGCAGCTACCCAGCAGCAG 193

RESULT 6

BY727796/c

LOCUS

DEFINITION BY727796 RIKEN full-length enriched, 10 days neonate medulla
oblongata Mus musculus cDNA clone B830037L10 5', mRNA sequence.

ACCESSION BY727796

VERSION BY727796.1 GI:27140923

KEYWORDS

SOURCE EST.

ORGANISM

Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 631)

REFERENCE

AUTHORS

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamana, A., Schonbach, C.,
Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Hume, D.A.,
Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A.,
Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S.,
Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani
A., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest
L.E., Gough, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.,
Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J.,
Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M.,
King, B.L., Kongaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons
P.B., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki
H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G.,
Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D.,
Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring
B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Secou
M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale
R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y.,
Watanabe, Y., Wells, C., Wilming, L.G., Wyszewski, B., Yanagisawa
M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,
Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura
M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii
Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata
K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander
E.S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

Email: genome-res@gsr.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda
S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K.,
Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno
H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K.,
Numazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H.,
Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y.,

KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Mus musculus

REFERENCE
1 (bases 1 to 743)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. NIH-MGC http://mgc.nci.nih.gov/

AUTHORS
TITLE
JOURNAL
COMMENT
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES
source
Location/Qualifiers
1..743
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5684935"
/tissue_type="whole brain"
/dev_stage="embryo 18.5 dpc"
/lab_host="DH10B (TI phage resistant)"
/clone_lib="NIH_BMAP_ECO"
/notes="Organ: Brain; Vector: pYX-Asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CAGCCAGCAGC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP). 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT
109 a 280 c 227 g 126 t 1 others

Alignment Scores:
Pred. No.: 343 Length: 743
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.77% Indels: 0
DB: 12 Gaps: 0

US-09-638-693-36 (1-133) x BM951073 (1-743)

Qy 104 LeuLeuGlnArgAlaThrGlnGlnGln 112
|||||
Db 485 CTCTGCGAGGAGCTACCCAGCAGCAG 459

RESULT 9
BU707783/c
LOCUS
DEFINITION
IMAGE: 6415190 5', mRNA sequence.
BU707783
758 bp mRNA linear EST 09-OCT-2002
BU707783
UT-M-FRO-cao-i-15-0-UT.r1 NIH_BMAP_FRO Mus musculus cDNA clone

VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Mus musculus

REFERENCE
1 (bases 1 to 758)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. NIH-MGC http://mgc.nci.nih.gov/

AUTHORS
TITLE
JOURNAL
COMMENT
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES
source
Location/Qualifiers
1..758
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE: 6415190"
/tissue_type="whole brain"
/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
/lab_host="DH10B (TI phage resistant)"
/clone_lib="NIH_BMAP_FRO"
/notes="Organ: Brain; Vector: pYX-Asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCAGCAGCAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT
114 a 286 c 240 g 118 t

Alignment Scores:
Pred. No.: 350 Length: 758
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.77% Indels: 0
DB: 13 Gaps: 0

US-09-638-693-36 (1-133) x BU707783 (1-758)

Qy 104 LeuLeuGlnArgAlaThrGlnGlnGln 112
|||||
Db 324 CTCTGCGAGGAGCTACCCAGCAGCAG 298

RESULT 10
BI968400/c
LOCUS
DEFINITION
IMAGE: 6415190 5', mRNA sequence.
BI968400
799 bp mRNA linear EST 23-OCT-2001
GM830005A12E12 Gm-r1083 Glycine max cDNA clone Gm-r1083-1751 3', mRNA sequence.

ACCESSION BI968400
 VERSION BI968400.1 GI:16342805
 EST.
 SOURCE Glycine max (soybean)
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 1 (bases 1 to 799)
 REFERENCE Vodkin, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Coryell, V.,
 Erpelzing, J., Rapp, C., Shoop, E., Fardinas, J., Liu, L. and Lewin, H.
 A Functional Genomics Program for Soybean (NSF 9872565)
 UNPUBLISHED
 TITLE Other ESTs: AW132308 corresponding to Gm-cl013-2340 (5')
 COMMENT Contact: Vodkin, L.O., PI, A Functional Genomics Program for
 Soybean (NSF 9872565)
 Lewin, H. A., Director, Keck Center for Comparative and Functional
 Genomics
 University of Illinois
 Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
 Tel: (217) 244-6147
 Fax: (217) 333-4582
 Email: l-vodkin@uiuc.edu
 This clone is available through: Incyte Genomics, 4633 World
 Parkway Circle St. Louis, Missouri 63134. Phone (800) 430-0030 or
 (314) 427-3222 FAX: (314) 427-3324. Web site:
 http://www.incyte.com/reagents/catalog.jsp?page=clones/collaboratio
 n/index
 Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.

FEATURES

source
 1. .799
 /organism="Glycine max"
 /mol_type="mRNA"
 /db_xref="taxon:3847"
 /clone="Gm-r1083-1751"
 /clone_lib="Gm-r1083"
 /note="The library Gm-r1083 is a sequence-driven, reracked
 set of 4,992 clones selected from cDNA libraries from
 various tissues and stages of development of soybean. It
 represents 1117 sequences from the progenitor library
 Gm-cl009 (from mature roots of 2 month old greenhouse
 grown 'Williams' soybean plants); 820 sequences from the
 progenitor library Gm-cl013 (from 2 to 3 week old whole
 plants of Williams); and 3055 sequences from library
 Gm-cl028 (from 'Supernod' plants whose seedlings were
 inoculated with Bradyrhizobium japonicum, courtesy of Dr.
 Gary Stacey). The 5' ESTs of the source clones from the
 different progenitor libraries was used to select
 singletons, or a representative of each contig, which were
 reracked to form library Gm-r1083. The cDNA clones of the
 reracked Gm-r1083 library were then sequenced at the 3'
 end. The contig analysis to select unique genes was
 performed by the laboratory of Ernest Retzel, Center for
 Computational Genomics and Bioinformatics, University of
 Minnesota, http://web.ahc.umn.edu/biodata/nsfsoy/.
 Reracking was performed by Incyte Genomics, St. Louis,
 http://www.incyte.com, and 3' sequencing by the Keck
 Center for Comparative and Functional Genomics, University
 of Illinois, http://www.lie.uiuc.edu/biotech/keck.html.
 Note: The corresponding 5' EST from each clone in the
 Gm-r1083 library is listed in the 'OTHER EST' field. The
 detailed information on the source library for each clone
 can also be obtained by referring to the Incyte Genomics
 clone ID of the original cDNA library that is also listed
 under 'OTHER EST'."

BASE COUNT 176 a 205 c 231 g 176 t 11 others
 ORIGIN

Alignment Scores:
 Pred. No.: 370 Length: 799
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 6.77% Indels: 0
 DB: 12 Gaps: 0
 US-09-638-693-36 (1-133) x BI968400 (1-799)
 QY 25 ThrThrSerThrTrpValLeuLeuGly 33
 |||||
 Db 234 ACAACTTCGACATGGGTTCTTCTTGGC 208
 |||||
 RESULT 11
 BX425691/c
 LOCUS BX425691 Homo sapiens NEUROBLASTOMA linear EST 15-MAY-2003
 DEFINITION CL0BB0262B08 5-PRIME, mRNA sequence.
 ACCESSION BX425691
 VERSION BX425691.1 GI:30774485
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 889)
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 TITLE Full-length cDNA libraries and normalization
 JOURNAL Unpublished
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 5578.f For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CL0BB0262B08RP1&cluster=5578.f. Contact :
 Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CL0BB0262B08RP1.

FEATURES

source
 1. .889
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CL0BB0262B08"
 /tissue_type="NEUROBLASTOMA"
 /clone_lib="Homo sapiens NEUROBLASTOMA"
 /note="Vector: pCMVSPORT.6; 1st strand cDNA was primed
 with a NotI-oligo(dT) primer. Five prime end enriched,
 double-strand cDNA was digested with Not I and cloned into
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.
 Library was not normalized."

BASE COUNT 273 a 231 c 215 g 131 t 39 others
 ORIGIN

Alignment Scores:

Pred. No.: 413 Length: 889
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 6.77% Indels: 0
 DB: 13 Gaps: 0

US-09-638-693-36 (1-133) x BX425691 (1-889)

QY 32 LeuGlyGlyValLeuAlaAlaLeuAla 40
 |||||

Db 547 TTAGGAGGAGTTTGGCTGCTTGGCT 521
 |||||

RESULT 12

BQ715452/c
 LOCUS BQ715452 904 bp mRNA linear EST 16-JUL-2002
 DEFINITION AGENCOURT_8290831 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6309329
 5', mRNA sequence.

ACCESSION BQ715452
 VERSION BQ715452.1 GI:21854351
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 904)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Susan L. Sullivan, PhD.
 cDNA Library Preparation: ResGen, Invitrogen Corp
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LLAM13729 row: n column: 18
 High quality sequence start: 17
 High quality sequence stop: 587.

FEATURES
 source
 Location/Qualifiers
 1..904
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6309329"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_129"
 /note="Organ: olfactory epithelium; Vector: pCMV-SPORT6.1.cdb; Site_1: EcorV; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."
 BASE COUNT 135 a 343 c 292 g 134 t
 ORIGIN

Alignment Scores:
 Pred. No.: 420 Length: 904
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 6.77% Indels: 0
 DB: 13 Gaps: 0

US-09-638-693-36 (1-133) x BQ715452 (1-904)

Qy 104 LeuLeuGlnArgAlaThrGlnGlnGln 112
 |||||
 Db 587 CTCCTGCAGCAGCTACCCAGCAG 561

RESULT 13
 BQ900568/c
 LOCUS BQ900568 1016 bp mRNA linear EST 16-AUG-2002
 DEFINITION AGENCOURT_8767827 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6314266 5', mRNA sequence.
 ACCESSION BQ900568
 VERSION BQ900568.1 GI:22292589
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 1016)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Susan L. Sullivan, PhD.
 cDNA Library Preparation: ResGen, Invitrogen Corp

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LLAM13742 row: 1 column: 11
 High quality sequence stop: 562.

FEATURES
 source
 Location/Qualifiers
 1..1016
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6314266"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_129"
 /note="Organ: olfactory epithelium; Vector: pCMV-SPORT6.1.cdb; Site_1: EcorV; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."
 BASE COUNT 149 a 376 c 329 g 162 t
 ORIGIN

Alignment Scores:
 Pred. No.: 475 Length: 1016
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 6.77% Indels: 0
 DB: 13 Gaps: 0

US-09-638-693-36 (1-133) x BQ900568 (1-1016)

Qy 104 LeuLeuGlnArgAlaThrGlnGlnGln 112
 |||||
 Db 543 CTCCTGCAGCAGCTACCCAGCAG 517

RESULT 14
 CNS05GS/c
 LOCUS CNS05GS 1101 bp DNA linear GSS 01-SEP-2000
 DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 001A17 of library B from Tetraodon nigroviridis, genomic survey sequence.
 ACCESSION AL335881.1 GI:8229639
 VERSION AL335881
 KEYWORDS GSS: genome survey sequence.
 SOURCE Tetraodon nigroviridis
 ORGANISM Tetraodon nigroviridis

REFERENCE 1
 AUTHORS Roest Crolius.H., Jaillon.O., Dasilva.C., Bouneau.L., Fisher.C., Bernot.A., Fizames.C., Wincker.P., Brottier.P., Quetier.F., Saurin.W. and Weissenbach.J.
 TITLE Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
 JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
 MEDLINE 20296633
 PUBMED 10835645
 REFERENCE 2

REFERENCE 1
 AUTHORS Roest Crolius.H., Jaillon.O., Dasilva.C., Ozouf-Costaz.C., Fizames.C., Fischer.C., Bouneau.L., Billault.A., Quetier.F., Saurin.W., Bernot.A. and Weissenbach.J.
 TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
 JOURNAL Genome Res. 10 (7), 939-949 (2000)
 MEDLINE 20359837
 PUBMED 10899143
 REFERENCE 3 (bases 1 to 1101)
 AUTHORS Genoscope.
 TITLE Direct Submission

JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr
- Web : www.genoscope.cns.fr)

COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES Location/Qualifiers
Source 1..1101
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="001A17"
/clone_lib="B"

BASE COUNT 282 a 243 c 285 g 283 t 8 others
ORIGIN /note="Genoscope sequence ID : C0TB001AA09C1-end : T7"

Alignment Scores: 517 Length: 1101
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.77% Indels: 0
DB: 29 Gaps: 0

US-09-638-693-36 (1-133) x CNS05G5S (1-1101)

QY 32 LeuGlyGlyValLeuAlaAlaLeuAla 40
|||||
Db 826 CTCGGCGGTGTTTGGCAGCTCTGGCT 800

RESULT 15

CD496483/C
LOCUS CD496483 1142 bp mRNA linear EST 12-JUN-2003
DEFINITION CDA22-E10.xld-t SHGC-CDA Gasterosteus aculeatus cDNA clone
CDA22-E10 5', mRNA sequence.

ACCESSION CD496483
VERSION CD496483.1 GI:31423514
KEYWORDS EST.

SOURCE Gasterosteus aculeatus (three spined stickleback)

ORGANISM Gasterosteus aculeatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
Gasterosteidae; Gasterosteus.

REFERENCE 1 (bases 1 to 1142)
Kingsley,D.M., Peichel,C., Balabhadra,S., Grimwood,J., Dickson,M.,
Schmutz,J. and Myers,R.M.

AUTHORS Expressed sequence tags from Gasterosteus aculeatus
Unpublished

TITLE Contact: Kingsley, DM

JOURNAL HHMI and Department of Developmental Biology
Stanford University School of Medicine
Beckman Center B300, 279 Campus Drive, Stanford, CA 94305-5329, USA
Tel: 650 725 5954
Fax: 650 725 7739

EMAIL Email: kingsley@cmgm.stanford.edu
Plate: 22
High quality sequence stop: 807.

FEATURES Location/Qualifiers
Source 1..1142

/organism="Gasterosteus aculeatus"
/mol_type="mRNA"
/strain="Salinas river, CA"
/db_xref="taxon:69293"
/clone="CDA22-E10"
/sex="mixed male and female"
/tissue_type="heads and internal organs combined"
/dev_stage="adult"
/clone_lib="SHGC-CDA"
/note="Vector: lambda ZAP Express/pBK-CMV; Site.1: EcoRI
(5' adaptor); Site.2: XhoI (3' linker primer); The mixed

organ cDNA library was generated using the ZAP-cDNA method
by Stratagene. First strand cDNA synthesis was primed with
a 50 bp linker primer containing an oligo dt sequence
preceded by a synthetic XhoI site. 5 prime adaptors were
used containing an EcoRI cohesive end. The finished cDNAs
were inserted in to the ZAP express vector
unidirectionally in the sense orientation with respect to
the lacZ promoter of pBK-CMV. An amplified library was
prepared from approximately 3 million primary clones in
the lambda ZAP Express vector. In vivo excision was then
used to generate individual pBK-CMV phagemid clones for
EST sequencing."

BASE COUNT 285 a 344 c 320 g 193 t
ORIGIN

Alignment Scores: 537 Length: 1142
Pred. No.: 9.00 Matches: 9
Score: 100.00% Conservatives: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 6.77% Gaps: 0
DB: 14

US-09-638-693-36 (1-133) x CD496483 (1-1142)

QY 32 LeuGlyGlyValLeuAlaAlaLeuAla 40
|||||

Db 117 CTGGAGGGGTCTTGTGCTGGCTGGCT 91

Search completed: August 29, 2003, 14:23:20
Job time : 1829 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 08:16:39 ; Search time 49 seconds
(without alignments)
430.829 Million cell updates/sec

Title: US-09-638-693-36
Perfect score: 699
Sequence: 1 QNEICLTHPIKTYIMACMSA.....VIEPIVTTNWKLEAFWHKH 133

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*

1:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	699	100.0	133	15 AAR63288	Polypeptide encode
2	691	98.9	133	15 AAR63289	Polypeptide encode
3	691	98.9	133	15 AAR63290	Polypeptide encode
4	691	98.9	209	15 AAR63390	HCV polypeptide se
5	672	96.1	133	15 AAR63286	Polypeptide encode
6	672	96.1	133	15 AAR63287	Polypeptide encode
7	652	93.3	128	14 AAR37937	HCV NS4 region con
8	648	92.7	128	14 AAR37932	HCV NS4 region fro
9	645	92.3	127	14 AAR37933	HCV NS4 region fro

10	645	92.3	127	14	AAR37936	HCV NS4 region fro
11	605	86.6	3023	17	AAR94462	Hepatitis C virus
12	590	84.4	117	14	AAR37934	HCV NS4 region fro
13	583	83.4	117	14	AAR37935	HCV NS4 region fro
14	539	77.1	3011	19	AAR77397	Hepatitis C virus
15	539	77.1	3011	24	ABP71460	Amino acid sequenc
16	539	77.1	3012	23	AAU99289	Hepatitis C virus
17	534	76.4	237	10	AAU90138	Peptide encoded by
18	534	76.4	363	13	AAR23999	Open reading frame
19	534	76.4	363	17	AAR90933	HCV antigen C100.
20	534	76.4	382	10	AAU90182	C terminus of supe
21	534	76.4	382	10	AAU92048	Sequence of the ca
22	534	76.4	460	10	AAU90141	protein sequence o
23	534	76.4	460	10	AAU92024	Polypeptide encode
24	534	76.4	592	14	AAR33565	CKS-HCV antigen fu
25	534	76.4	592	22	AAU69023	HCV recombinant an
26	534	76.4	594	14	AAR33566	CKS-HCV antigen fu
27	534	76.4	594	22	AAU69024	HCV recombinant an
28	534	76.4	597	13	AAR21571	HCV CKS-C100D1 - p
29	534	76.4	597	14	AAR33638	HCV C100D1 recombi
30	534	76.4	597	14	AAR33580	HCV C100D1 recombi
31	534	76.4	597	14	AAR33600	HCV C100D1 recombi
32	534	76.4	597	22	AAU51378	HCV recombinant an
33	534	76.4	599	13	AAR21572	HCV CKS-C100D2 - p
34	534	76.4	599	14	AAR33639	HCV C100D2 recombi
35	534	76.4	599	14	AAR33581	HCV C100D2 recombi
36	534	76.4	599	14	AAR33601	HCV C100D2 recombi
37	534	76.4	599	22	AAU51379	HCV recombinant an
38	534	76.4	613	14	AAR33567	CKS-HCV antigen fu
39	534	76.4	613	22	AAU69025	HCV recombinant an
40	534	76.4	858	10	AAU90146	ORF extending thro
41	534	76.4	859	10	AAU92029	Sequence encoded i
42	534	76.4	971	22	AAU69027	HCV recombinant an
43	534	76.4	973	14	AAR33570	CKS-HCV antigen fu
44	534	76.4	973	22	AAU69028	HCV recombinant an
45	534	76.4	992	14	AAR33571	CKS-HCV antigen fu

ALIGNMENTS

RESULT 1
AAR63288
ID AAR63288 standard; Protein; 133 AA.
XX
AC AAR63288;
XX
25-MAR-2003 (updated)
DT 01-AUG-1995 (first entry)
DT
XX
XX Polypeptide encoded by hepatitis C virus NS3/NS4 sequence.
DE
XX Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
KW Classification; immunisation; prophylaxis; serotyping.
XX
XX Hepatitis C virus type 3.
OS
XX
XX W09425601-A2.
PN
XX
PD
XX
XX 10-NOV-1994.
XX
XX 27-APR-1994; 94WO-EP01323.
PF
XX
XX 27-APR-1993; 93EP-0401099.
PR
XX
XX 05-AUG-1993; 93EP-0402019.
PR
XX
XX (INNO-) INNOGENETICS NV SA.
PA
XX
XX Maertens G, Stuyver L;
PI
XX
XX WPI; 1994-358277/44.
DR
XX
XX N-PSDB; AAQ78040.

PT New polynucleotide sequences from hepatitis C virus - and related
PT vectors, polypeptide(s) and antibodies, useful for immunisation,
PS treatment, diagnosis and typing of HCV isolates
XX Claim 11; Page 125; 404pp; English.
XX Compositions comprising at least 5, and pref. 8 or more contiguous
CC nucleotides selected from an HCV type 3 genomic sequence, more
CC particularly (i) the region spanning positions 417-957 of the
CC Core/E1 region of HCV subtype 3a; (ii) the region spanning positions
CC 4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
CC positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
CC region spanning positions 8023-8235 of the NS5 region of the BR36
CC subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
CC sequence, may be used as primers to amplify nucleic acid from an
CC isolate belonging to a specific genotype, or as a probe for specific
CC detection/classification of nucleic acid. Polypeptides encoded by
CC the nucleotides in such compositions may be used for immunisation
CC against HCV, for the detection of antibodies directed against HCV
CC and for serotyping. This sequence corresponds to the NS3/NS4
CC region of HCV subtype 3a and is taken from a clone designated
CC BR36-20-164.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX Sequence 133 AA;
SQ Query Match 100.0%; Score 699; DB 15; Length 133;
Best Local Similarity 100.0%; Pred. No. 1.4e-73;
Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 QNEICLTHTPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
Db 1 QNEICLTHTPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
Qy 61 PAIVPDKEVLYQQYDEMEECSSAAPYTEQAQVIAHOFKGVGLGLLQRTQQQAVIEPIVT 120
Db 61 PAIVPDKEVLYQQYDEMEECSSAAPYTEQAQVIAHOFKGVGLGLLQRTQQQAVIEPIVT 120
Qy 121 TNWOKLEAFWHKH 133
Db 121 TNWOKLEAFWHKH 133
RESULT 2
AAR63289
ID AAR63289 standard; Protein; 133 AA.
XX AAR63289;
XX 25-MAR-2003 (updated)
DT 01-AUG-1995 (first entry)
XX Polypeptide encoded by hepatitis C virus NS3/NS4 sequence.
XX Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
KW classification; immunisation; prophylaxis; serotyping.
XX Hepatitis C virus type 3.
OS WO9425601-A2.
XX 10-NOV-1994.
XX 27-APR-1994; 94WO-EP01323.
XX 27-APR-1993; 93EP-0401099.
PR 05-AUG-1993; 93EP-0402019.
XX (INNO-) INNOGENETICS NV SA.
XX Maertens G, Stuyver L;
XX WPI; 1994-358277/44.
DR

DR N-PSDB; AAQ78041.
XX New polynucleotide sequences from hepatitis C virus - and related
PT vectors, polypeptide(s) and antibodies, useful for immunisation,
PT treatment, diagnosis and typing of HCV isolates
XX Claim 11; Page 127; 404pp; English.
XX Compositions comprising at least 5, and pref. 8 or more contiguous
CC nucleotides selected from an HCV type 3 genomic sequence, more
CC particularly (i) the region spanning positions 417-957 of the
CC Core/E1 region of HCV subtype 3a; (ii) the region spanning positions
CC 4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
CC positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
CC region spanning positions 8023-8235 of the NS5 region of the BR36
CC subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
CC sequence, may be used as primers to amplify nucleic acid from an
CC isolate belonging to a specific genotype, or as a probe for specific
CC detection/classification of nucleic acid. Polypeptides encoded by
CC the nucleotides in such compositions may be used for immunisation
CC against HCV, for the detection of antibodies directed against HCV
CC and for serotyping. This sequence corresponds to the NS3/NS4
CC region of HCV subtype 3a and is taken from a clone designated
CC BR36-20-166.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX Sequence 133 AA;
SQ Query Match 98.9%; Score 691; DB 15; Length 133;
Best Local Similarity 99.2%; Pred. No. 1.2e-72;
Matches 132; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 QNEICLTHTPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
Db 1 QNEICLTHTPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
Qy 61 PAIVPDKEVLYQQYDEMEECSSAAPYTEQAQVIAHOFKGVGLGLLQRTQQQAVIEPIVT 120
Db 61 PAIVPDKEVLYQQYDEMEECSSAAPYTEQAQVIAHOFKGVGLGLLQRTQQQAVIEPIVT 120
Qy 121 TNWOKLEAFWHKH 133
Db 121 TNWOKLEAFWHKH 133
RESULT 3
AAR63290
ID AAR63290 standard; Protein; 133 AA.
XX AAR63290;
XX 25-MAR-2003 (updated)
DT 01-AUG-1995 (first entry)
XX Polypeptide encoded by hepatitis C virus NS3/NS4 sequence.
XX Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
KW classification; immunisation; prophylaxis; serotyping.
XX Hepatitis C virus type 3.
OS WO9425601-A2.
XX 10-NOV-1994.
XX 27-APR-1994; 94WO-EP01323.
XX 27-APR-1993; 93EP-0401099.
PR 05-AUG-1993; 93EP-0402019.
XX (INNO-) INNOGENETICS NV SA.
XX Maertens G, Stuyver L;
XX WPI; 1994-358277/44.
DR

```

XX WPI; 1994-358277/44.
DR N-PSDB; AAQ78042.
XX
PT New polynucleotide sequences from hepatitis C virus - and related
PT vectors, polypeptide(s) and antibodies, useful for immunisation,
PT treatment, diagnosis and typing of HCV isolates
XX
PS Clalm 11; Page 138-129; 404pp; English.
XX
CC Compositions comprising at least 5, and pref. 8 or more contiguous
CC nucleotides selected from an HCV type 3 genomic sequence, more
CC particularly (i) the region spanning positions 417-957 of the
CC Core/EL region of HCV subtype 3a; (ii) the region spanning positions
CC 4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
CC positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
CC region spanning positions 8023-8235 of the NS5 region of the BR36
CC subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
CC sequence, may be used as primers to amplify nucleic acid from an
CC isolate belonging to a specific genotype, or as a probe for specific
CC detection/classification of nucleic acid. Polypeptides encoded by
CC the nucleotides in such compositions may be used for immunisation
CC against HCV, for the detection of antibodies directed against HCV
CC and for serotyping. This sequence corresponds to the NS3/NS4
CC region of HCV subtype 3a and is taken from a clone designated
CC BR36-20-165.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 133 AA;
Query Match 98.9%; Score 691; DB 15; Length 133;
Best Local Similarity 99.2%; Pred. No. 1.2e-72;
Matches 132; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ONEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSVGCVVIVGHTELGGK 60
DB 1 ONEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSVGCVVIVGHTELGGK 60
QY 61 PAIVDPKEVLQOYDEMEECQAAPYIEQAQVIAHOFKVKVGLLQRLATQQQAVIEPIVT 120
DB 61 PAIVDPKEVLQOYDEMEECQAAPYIEQAQVIAHOFKVKVGLLQRLATQQQAVIEPIVT 120
QY 121 TNWQKLEAFWHKH 133
DB 121 TNWQKLEAFWHKH 133
RESULT 4
AAR63390
ID AAR63390 standard; Protein; 209 AA.
XX
AC AAR63390;
XX
DT 25-MAR-2003 (updated)
DT 18-AUG-1995 (first entry)
XX
DE HCV polypeptide sequence.
XX
KW Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
KW classification; immunisation; prophylaxis; serotyping.
XX
OS Hepatitis C virus.
XX
PN WO9425601-A2.
XX
PD 10-NOV-1994.
XX
PF 27-APR-1994; 94WO-EP01323.
XX
PR 27-APR-1993; 93EP-0401099.
PR 05-AUG-1993; 93EP-0402019.
XX
PA (INNO-) INNOGENETICS NV SA.

```

```

XX Maertens G, Stuyver L;
PI
XX WPI; 1994-358277/44.
DR N-PSDB; AAQ78125.
XX
PT New polynucleotide sequences from hepatitis C virus - and related
PT vectors, polypeptide(s) and antibodies, useful for immunisation,
PT treatment, diagnosis and typing of HCV isolates
XX
PS Disclosure; Page 274-275; 404pp; English.
XX
CC Compositions comprising at least 5, and pref. 8 or more contiguous
CC nucleotides selected from an HCV type 3 genomic sequence, more
CC particularly (i) the region spanning positions 417-957 of the
CC Core/EL region of HCV subtype 3a; (ii) the region spanning positions
CC 4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
CC positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
CC region spanning positions 8023-8235 of the NS5 region of the BR36
CC subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
CC sequence, may be used as primers to amplify nucleic acid from an
CC isolate belonging to a specific genotype, or as a probe for specific
CC detection/classification of nucleic acid. Polypeptides encoded by
CC the nucleotides in such compositions may be used for immunisation
CC against HCV, for the detection of antibodies directed against HCV
CC and for serotyping.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 209 AA;
Query Match 98.9%; Score 691; DB 15; Length 209;
Best Local Similarity 99.2%; Pred. No. 2.2e-72;
Matches 132; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ONEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSVGCVVIVGHTELGGK 60
DB 77 ONEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSVGCVVIVGHTELGGK 136
QY 61 PAIVDPKEVLQOYDEMEECQAAPYIEQAQVIAHOFKVKVGLLQRLATQQQAVIEPIVT 120
DB 137 PAIVDPKEVLQOYDEMEECQAAPYIEQAQVIAHOFKVKVGLLQRLATQQQAVIEPIVT 196
QY 121 TNWQKLEAFWHKH 133
DB 197 TNWQKLEAFWHKH 209
RESULT 5
AAR63286
ID AAR63286 standard; Protein; 133 AA.
XX
AC AAR63286;
XX
DT 25-MAR-2003 (updated)
DT 01-AUG-1995 (first entry)
XX
DE Polypeptide encoded by hepatitis C virus NS3/NS4 sequence.
XX
KW Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
KW classification; immunisation; prophylaxis; serotyping.
XX
OS Hepatitis C virus type 3.
XX
PN WO9425601-A2.
XX
PD 10-NOV-1994.
XX
PF 27-APR-1994; 94WO-EP01323.
XX
PR 27-APR-1993; 93EP-0401099.
PR 05-AUG-1993; 93EP-0402019.
XX
PA (INNO-) INNOGENETICS NV SA.

```

```

XX PI Maertens G, Stuyver L;
XX WPI; 1994-358277/44.
XX DR N-PSDB; RAQ78038.
XX PT New polynucleotide sequences from hepatitis C virus - and related
XX PT vectors, polypeptide(s) and antibodies, useful for immunisation,
XX PT treatment, diagnosis and typing of HCV isolates
XX PS Claim 11; Page 121-122; 404pp; English.
XX CC Compositions comprising at least 5, and pref. 8 or more contiguous
XX CC nucleotides selected from an HCV type 3 genomic sequence, more
XX CC particularly (i) the region spanning positions 417-957 of the
XX CC Core/E1 region of HCV subtype 3a; (ii) the region spanning positions
XX CC 4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
XX CC positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
XX CC region spanning positions 8023-8235 of the NS5 region of the BR36
XX CC subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
XX CC sequence, may be used as primers to amplify nucleic acid from an
XX CC isolate belonging to a specific genotype, or as a probe for specific
XX CC detection/classification of nucleic acid. Polypeptides encoded by
XX CC the nucleotides in such compositions may be used for immunisation
XX CC against HCV, for the detection of antibodies directed against HCV
XX CC and for serotyping. This sequence corresponds to the NS3/NS4
XX CC region of HCV subtype 3a and is taken from a clone designated
XX CC HD10-1-25.
XX CC (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 133 AA;

Query Match 96.1%; Score 672; DB 15; Length 133;
Best Local Similarity 93.2%; Pred. No. 2e-70;
Matches 124; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 ONEICLTHTPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGK 60
DB 1 ONEICLTHTPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGK 60
QY 61 PALVPDKEVLYQOYDEMECSQAAPYIEQAQVIAHOFKGVGLLQORATQQQAVIEPIVT 120
DB 61 PALVPDKEVLYQOYDEMECSQAAPYIEQAQVIAHOFKGVGLLQORATQQQAVIEPIVT 120
QY 121 TNWQKLEAFWHKH 133
DB 121 SNWQKLETFWHKH 133

RESULT 6
AAR63287
ID AAR63287 standard; Protein; 133 AA.
XX AC AAR63287;
XX DT 25-MAR-2003 (updated)
XX DT 01-AUG-1995 (first entry)
XX DE Polypeptide encoded by hepatitis C virus NS3/NS4 sequence.
XX KW Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
XX KW classification; immunisation; prophylaxis; serotyping.
XX OS Hepatitis C virus type 3.
XX PN WO9425601-A2.
XX PD 10-NOV-1994.
XX PF 27-APR-1994; 94WO-EP01323.
XX PR 27-APR-1993; 93EP-0401099.
XX PR 05-AUG-1993; 93EP-0402019.

```

```

XX PA (INNO-) INNOGENETICS NV SA.
XX PI Maertens G, Stuyver L;
XX DR WPI; 1994-358277/44.
XX DR N-PSDB; RAQ78039.
XX PT New polynucleotide sequences from hepatitis C virus - and related
XX PT vectors, polypeptide(s) and antibodies, useful for immunisation,
XX PT treatment, diagnosis and typing of HCV isolates
XX PS Claim 11; Page 123-124; 404pp; English.
XX CC Compositions comprising at least 5, and pref. 8 or more contiguous
XX CC nucleotides selected from an HCV type 3 genomic sequence, more
XX CC particularly (i) the region spanning positions 417-957 of the
XX CC Core/E1 region of HCV subtype 3a; (ii) the region spanning positions
XX CC 4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
XX CC positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
XX CC region spanning positions 8023-8235 of the NS5 region of the BR36
XX CC subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
XX CC sequence, may be used as primers to amplify nucleic acid from an
XX CC isolate belonging to a specific genotype, or as a probe for specific
XX CC detection/classification of nucleic acid. Polypeptides encoded by
XX CC the nucleotides in such compositions may be used for immunisation
XX CC against HCV, for the detection of antibodies directed against HCV
XX CC and for serotyping. This sequence corresponds to the NS3/NS4
XX CC region of HCV subtype 3a and is taken from a clone designated
XX CC HD10-1-3.
XX CC (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 133 AA;

Query Match 96.1%; Score 672; DB 15; Length 133;
Best Local Similarity 93.2%; Pred. No. 2e-70;
Matches 124; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 ONEICLTHTPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGK 60
DB 1 ONEICLTHTPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGK 60
QY 61 PALVPDKEVLYQOYDEMECSQAAPYIEQAQVIAHOFKGVGLLQORATQQQAVIEPIVT 120
DB 61 PALVPDKEVLYQOYDEMECSQAAPYIEQAQVIAHOFKGVGLLQORATQQQAVIEPIVT 120
QY 121 TNWQKLEAFWHKH 133
DB 121 SNWQKLETFWHKH 133

RESULT 7
AAR37937
ID AAR37937 standard; Protein; 128 AA.
XX AC AAR37937;
XX DT 25-MAR-2003 (updated)
XX DT 23-SEP-1993 (first entry)
XX DE HCV NS4 region consensus sequence.
XX KW Non-coding region; hepatitis C virus; blood donor; type 2; type 1;
XX KW HCV; NS-5; phylogeny; differentiation; NS-3; core region; type 3.
XX OS Synthetic.
XX PN WO9310239-A2.
XX PD 27-MAY-1993.
XX PF 20-NOV-1992; 92WO-GB02143.

```


PR 21-NOV-1991; 91GB-0024696.
 PR 24-JUN-1992; 92GB-0013362.
 PA (COMM-) COMMON SERVICES AGENCY.
 XX

PI Chan S, Simmonds P, Yap PL;
 XX

DR WPI; 1993-182554/22.
 DR N-PSDB; AAQ43111.

XX DNA encoding antigenic peptide(s) of new types of hepatitis C
 PT virus - for diagnosing and treating HCV infection, screening
 PT blood samples and identifying different HCV types
 XX

PS Disclosure; Fig 9b; 120pp; English.
 XX

CC The sequences given in AAR37932-37 show amino acids 1638-1765 of the NS4
 CC region of hepatitis C virus-3 (HCV-3) samples from 5 blood donors
 CC and a consensus sequence. Analysis of this and other regions of the
 CC HCV genome revealed the existence of three distinct groups of HCV.
 CC Analysis of the region encompassing -255 to -62 of the 5' non coding
 CC region (see AAQ43058-75) showed a difference of 9-14% in the nucleotide
 CC sequences between the three groups. Two of the groups identified were
 CC similar to those of HCV variants termed type 1 and 2, whilst the third
 CC appeared to represent a novel type of virus. Comparison of the NS3
 CC region (see AAR37927-30) showed a high degree of sequence diversity with
 CC type 3 being phylogenetically different to type 1 and 2. The same
 CC degree of differentiation was noted in the NS-5 (see AAR37923-26) and core
 CC region between type 3 and type 1 sequences.
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 128 AA;
 SQ

Query Match 93.3%; Score 652; DB 14; Length 128;
 Best Local Similarity 97.6%; Pred. No. 4.2e-68;
 Matches 124; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 7 THPTKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSVGCVVIVGHIELGKPAIVPD 66
 |||||
 Db 1 THPTKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSVGCVVIVGHIELGKPAIVPD 60
 QY 67 KEVLVQQYDEMEECQAAPYIEQAQVIAHOFKGVGLLQRTAQQAQVIEPIVTNNQKL 126
 |||||
 Db 61 KEVLVQQYDEMEECQAAPYIEQAQVIAHOFKGVGLLQRTAQQAQVIEPIVTNNQKL 120
 QY 127 EAFWHKH 133
 |||||
 Db 121 EAFWHKH 127

RESULT 8
 AAR37932
 ID AAR37932 standard; Protein; 128 AA.
 XX

AC AAR37932;
 XX

DT 25-MAR-2003 (updated)
 DT 23-SEP-1993 (first entry)
 XX

DE HCV NS4 region from donor T0040.
 XX

XX Non-coding region; hepatitis C virus; blood donor; type 2; type 1;
 KW HCV; NS-5; phylogeny; differentiation; NS-3; core region; type 3.
 XX

OS Hepatitis C virus.
 XX

PN WO9310239-A2.
 XX

PD 27-MAY-1993.
 XX

PF 20-NOV-1992; 92WO-CB02143.
 XX

XX 21-NOV-1991; 91GB-0024696.
 PR

PR 24-JUN-1992; 92GB-0013362.
 XX (COMM-) COMMON SERVICES AGENCY.
 XX

PI Chan S, Simmonds P, Yap PL;
 XX

DR WPI; 1993-182554/22.
 DR N-PSDB; AAQ43106.

XX DNA encoding antigenic peptide(s) of new types of hepatitis C
 PT virus - for diagnosing and treating HCV infection, screening
 PT blood samples and identifying different HCV types
 XX

PS Disclosure; Fig 9b; 120pp; English.
 XX

CC The sequences given in AAR37932-37 show amino acids 1638-1765 of the NS4
 CC region of hepatitis C virus-3 (HCV-3) samples from 5 blood donors
 CC and a consensus sequence. Analysis of this and other regions of the
 CC HCV genome revealed the existence of three distinct groups of HCV.
 CC Analysis of the region encompassing -255 to -62 of the 5' non coding
 CC region (see AAQ43058-75) showed a difference of 9-14% in the nucleotide
 CC sequences between the three groups. Two of the groups identified were
 CC similar to those of HCV variants termed type 1 and 2, whilst the third
 CC appeared to represent a novel type of virus. Comparison of the NS3
 CC region (see AAR37927-30) showed a high degree of sequence diversity with
 CC type 3 being phylogenetically different to type 1 and 2. The same
 CC degree of differentiation was noted in the NS-5 (see AAR37923-26) and core
 CC region between type 3 and type 1 sequences.
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 128 AA;
 SQ

Query Match 92.7%; Score 648; DB 14; Length 128;
 Best Local Similarity 96.9%; Pred. No. 1.2e-67;
 Matches 123; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 7 THPTKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSVGCVVIVGHIELGKPAIVPD 66
 |||||
 Db 1 THPTKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSVGCVVIVGHIELGKPAIVPD 60
 QY 67 KEVLVQQYDEMEECQAAPYIEQAQVIAHOFKGVGLLQRTAQQAQVIEPIVTNNQKL 126
 |||||
 Db 61 KEVLVQQYDEMEECQAAPYIEQAQVIAHOFKGVGLLQRTAQQAQVIEPIVTNNQKL 120
 QY 127 EAFWHKH 133
 |||||
 Db 121 EAFWHKH 127

RESULT 9
 AAR37933
 ID AAR37933 standard; Protein; 127 AA.
 XX

AC AAR37933;
 XX

DT 25-MAR-2003 (updated)
 DT 23-SEP-1993 (first entry)
 XX

DE HCV NS4 region from donor T0038.
 XX

XX Non-coding region; hepatitis C virus; blood donor; type 2; type 1;
 KW HCV; NS-5; phylogeny; differentiation; NS-3; core region; type 3.
 XX

OS Hepatitis C virus.
 XX

PN WO9310239-A2.
 XX

PD 27-MAY-1993.
 XX

PF 20-NOV-1992; 92WO-CB02143.
 XX

XX 21-NOV-1991; 91GB-0024696.
 PR

PR 24-JUN-1992; 92GB-0013362.

```

XX (COMM-) COMMON SERVICES AGENCY.
PA Chan S, Simmonds P, Yap PL;
PI WPI; 1993-182554/22.
XX N-PSDB; AAQ43107.
XX DNA encoding antigenic peptide(s) of new types of hepatitis C
PT virus - for diagnosing and treating HCV infection, screening
PT blood samples and identifying different HCV types
XX Disclosure; Fig 9b; 120pp; English.
PS
XX The sequences given in AAR37932-37 show amino acids 1638-1765 of the NS4
CC region of hepatitis C virus-3 (HCV-3) samples from 5 blood donors
CC and a consensus sequence. Analysis of this and other regions of the
CC HCV genome revealed the existence of three distinct groups of HCV.
CC Analysis of the region encompassing -255 to -62 of the 5' non coding
CC region (see AAQ43058-75) showed a difference of 9-14% in the nucleotide
CC sequences between the three groups. Two of the groups identified were
CC similar to those of HCV variants termed type 1 and 2, whilst the third
CC appeared to represent a novel type of virus. Comparison of the NS3
CC region (see AAR37927-30) showed a high degree of sequence diversity with
CC type 3 being phylogenetically different to type 1 and 2. The same
CC degree differentiation was noted in the NS-5 (see AAR37923-26) and core
CC region between type 3 and type 1 sequences.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 127 AA;
Query Match 92.3%; Score 645; DB 14; Length 127;
Best Local Similarity 96.8%; Pred. No. 2.7e-67;
Matches 122; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 8 HPITKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGKPAIVPDK 67
DB 1 HPVTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGKPAIVPDK 60
QY 68 EVLYQQYDEMECSQAAPYIEQAQVIAHQFKVGLGLQRTAQQAQVIEPIVTTNWKLE 127
DB 61 EVLYQQYDEMECSQAAPYIEQAQVIAHQFKVGLGLQRTAQQAQVIEPIVTTNWKLE 120
QY 128 AFWKH 133
DB 121 AFWKH 126
RESULT 10
AAR37936
ID AAR37936 standard; Protein; 127 AA.
XX
AC AAR37936;
XX
XX 25-MAR-2003 (updated)
DT 23-SEP-1993 (first entry)
XX
DE HCV NS4 region from donor T1787.
XX
XX Non-coding region; hepatitis C virus; blood donor; type 2; type 1;
KW HCV; NS-5; phylogeny; differentiation; NS-3; core region; type 3.
XX
XX Hepatitis C virus.
OS
XX W09310239-A2.
XX
XX 27-MAY-1993.
PD
XX 20-NOV-1992; 92WO-GB02143.
XX
XX 21-NOV-1991; 91GB-0024596.
PR
XX 24-JUN-1992; 92GB-0013362.
XX

```

```

PA (COMM-) COMMON SERVICES AGENCY.
XX Chan S, Simmonds P, Yap PL;
XX WPI; 1993-182554/22.
DR N-PSDB; AAQ43110.
XX
XX DNA encoding antigenic peptide(s) of new types of hepatitis C
PT virus - for diagnosing and treating HCV infection, screening
PT blood samples and identifying different HCV types
XX Disclosure; Fig 9b; 120pp; English.
PS
XX The sequences given in AAR37932-37 show amino acids 1638-1765 of the NS4
CC region of hepatitis C virus-3 (HCV-3) samples from 5 blood donors
CC and a consensus sequence. Analysis of this and other regions of the
CC HCV genome revealed the existence of three distinct groups of HCV.
CC Analysis of the region encompassing -255 to -62 of the 5' non coding
CC region (see AAQ43058-75) showed a difference of 9-14% in the nucleotide
CC sequences between the three groups. Two of the groups identified were
CC similar to those of HCV variants termed type 1 and 2, whilst the third
CC appeared to represent a novel type of virus. Comparison of the NS3
CC region (see AAR37927-30) showed a high degree of sequence diversity with
CC type 3 being phylogenetically different to type 1 and 2. The same
CC degree differentiation was noted in the NS-5 (see AAR37923-26) and core
CC region between type 3 and type 1 sequences.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 127 AA;
Query Match 92.3%; Score 645; DB 14; Length 127;
Best Local Similarity 96.8%; Pred. No. 2.7e-67;
Matches 122; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 8 HPITKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGKPAIVPDK 67
DB 1 HPITKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGKPAIVPDK 60
QY 68 EVLYQQYDEMECSQAAPYIEQAQVIAHQFKVGLGLQRTAQQAQVIEPIVTTNWKLE 127
DB 61 EVLYQQYDEMECSQAAPYIEQAQVIAHQFKVGLGLQRTAQQAQVIEPIVTTNWKLE 120
QY 128 AFWKH 133
DB 121 AFWKH 126
RESULT 11
AAR94462
ID AAR94462 standard; Protein; 3023 AA.
XX
AC AAR94462;
XX
XX 20-SEP-1996 (first entry)
DT Hepatitis C virus polypeptide.
XX
XX hepatitis C virus; antibody; detection; diagnosis; vaccine;
KW classify; subtype.
XX
XX Hepatitis C virus.
OS
XX Key Location/Qualifiers
FH Peptide 1505..1520
FT /note="this part of the sequence is missing from
FT Peptide 2433..2448
FT /note="this part of the sequence is missing from
FT
XX JP08056672-A.
XX 05-MAR-1996.
PD

```

```

XX PF 26-AUG-1994; 94JP-0223933.
XX PR 26-AUG-1994; 94JP-0223933.
XX PA (SAYA/) SAYAMA K.
XX DR WPI; 1996-182301/19.
XX DR N-PSDB; AAT13279.
XX PT Hepatitis C virus genomic RNA, DNA and related proteins - useful for
XX PT detection, diagnosis and identification of hepatitis C virus
XX PT sub-type
XX PS Claim 4; Page 16-23; 25pp; Japanese.
XX CC The present sequence is a polypeptide comprising a 3023 amino acid
XX CC sequence which is encoded by a cDNA (AAT13279) to a hepatitis C virus
XX CC (HCV) genomic RNA. The protein can be easily detected by antibodies
XX CC in an assay for the detection of HCV. The DNA and the protein are
XX CC useful for classifying the subtype of HCV. At least a part of the
XX CC protein may be used as a vaccine against HCV.
XX SQ Sequence 3023 AA;
      Query Match 86.6%; Score 605; DB 17; Length 3023;
      Best Local Similarity 84.2%; Pred. No. 9.1e-61;
      Matches 112; Conservative 11; Mismatches 10; Indels 0; Gaps 0;
QY 1 QNEICLTHPTITRYIMACMSADLEVTSTWVLGGVLAALAAAYCLSGCVVIVGHIELGK 60
Db 1640 QNDICMTHPTITRYIMACMSADLEVTSTWVLGGVLAALAAAYCLSGCVVIVGHIELGK 1699
QY 61 PAIVDPKLVLYOQYDEMECSQAAPYIEQAQVIAHOFKGVLLGLLQRTQQQAVIEPIVT 120
Db 1700 PALVDPKLVLYOQYDEMECSQAAPYIEQAQVIAHOFKGVLLGLLQRTQQQAVIEPIVT 1759
QY 121 TNWQKLEAFWHKH 133
Db 1760 SQWQKAEAFWQH 1772
      RESULT 12
      AAR37934
      ID AAR37934 standard; Protein; 117 AA.
      AC AAR37934;
      XX
      DT 25-MAR-2003 (updated)
      DT 23-SEP-1993 (first entry)
      XX
      DE HCV NS4 region from donor T0036.
      XX
      KW Non-coding region; hepatitis C virus; blood donor; type 2; type 1;
      KW HCV; NS-5; phylogeny; differentiation; NS-3; core region; type 3.
      XX
      OS Hepatitis C virus.
      XX
      PN W09310239-A2.
      PD
      XX
      XX 27-MAY-1993.
      XX
      PF 20-NOV-1992; 92WO-GB02143.
      XX
      PR 21-NOV-1991; 91GB-0024696.
      PR 24-JUN-1992; 92GB-0013362.
      XX
      PA (COMM-) COMMON SERVICES AGENCY.
      XX
      PI Chan S, Simmonds P, Yap PL;
      XX WPI; 1993-182554/22.
      DR N-PSDB; AAQ43108.

```

```

XX PT DNA encoding antigenic peptide(s) of new types of hepatitis C
XX PT virus - for diagnosing and treating HCV infection, screening
XX PT blood samples and identifying different HCV types
XX PS Disclosure; Fig 9b; 120pp; English.
XX CC The sequences given in AAR37932-37 show amino acids 1638-1765 of the NS4
XX CC region of hepatitis C virus-3 (HCV-3) samples from 5 blood donors
XX CC and a consensus sequence. Analysis of this and other regions of the
XX CC HCV genome revealed the existence of three distinct groups of HCV.
XX CC Analysis of the region encompassing -255 to -62 of the 5' non coding
XX CC region (see AAQ43058-75) showed a difference of 9-14% in the nucleotide
XX CC sequences between the three groups. Two of the groups identified were
XX CC similar to those of HCV variants termed type 1 and 2, whilst the third
XX CC appeared to represent a novel type of virus. Comparison of the NS3
XX CC region (see AAR37927-30) showed a high degree of sequence diversity with
XX CC type 3 being phylogenetically different to type 1 and 2. The same
XX CC degree of differentiation was noted in the NS-5 (see AAR37923-26) and core
XX CC region between type 3 and type 1 sequences.
XX CC (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 117 AA;
      Query Match 84.4%; Score 590; DB 14; Length 117;
      Best Local Similarity 97.4%; Pred. No. 6.6e-61;
      Matches 114; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 8 HPITIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSGCVVIVGHIELGKPAIVDPK 67
Db 1 HPITIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSGCVVIVGHIELGKPAIVDPK 60
QY 68 EYLYOQYDEMECSQAAPYIEQAQVIAHOFKGVLLGLLQRTQQQAVIEPIVTNWQ 124
Db 61 EYLYOQYDEMECSQAAPYIEQAQVIAHOFKGVLLGLLQRTQQQAVIEPIVTNWQ 117
      RESULT 13
      AAR37935
      ID AAR37935 standard; Protein; 117 AA.
      AC AAR37935;
      XX
      DT 25-MAR-2003 (updated)
      DT 23-SEP-1993 (first entry)
      XX
      DE HCV NS4 region from donor T0026.
      XX
      KW Non-coding region; hepatitis C virus; blood donor; type 2; type 1;
      KW HCV; NS-5; phylogeny; differentiation; NS-3; core region; type 3.
      XX
      OS Hepatitis C virus.
      XX
      PN W09310239-A2.
      PD
      XX
      XX 27-MAY-1993.
      XX
      PF 20-NOV-1992; 92WO-GB02143.
      XX
      PR 21-NOV-1991; 91GB-0024696.
      PR 24-JUN-1992; 92GB-0013362.
      XX
      PA (COMM-) COMMON SERVICES AGENCY.
      XX
      PI Chan S, Simmonds P, Yap PL;
      XX WPI; 1993-182554/22.
      DR N-PSDB; AAQ43109.
      XX
      PT DNA encoding antigenic peptide(s) of new types of hepatitis C
      PT virus - for diagnosing and treating HCV infection, screening
      PT blood samples and identifying different HCV types
      XX

```

PS Disclosure; Fig 9b; 120pp; English.

XX The sequences given in AAR37932-37 show amino acids 1638-1765 of the NS4
CC region of hepatitis C virus-3 (HCV-3) samples from 5 blood donors
CC and a consensus sequence. Analysis of this and other regions of the
CC HCV genome revealed the existence of three distinct groups of HCV.
CC Analysis of the region encompassing -255 to -62 of the 5' non coding
CC region (see AAQ43058-75) showed a difference of 9-14% in the nucleotide
CC sequences between the three groups. Two of the groups identified were
CC similar to those of HCV variants termed type 1 and 2, whilst the third
CC appeared to represent a novel type of virus. Comparison of the NS3
CC region (see AAR37927-30) showed a high degree of sequence diversity with
CC type 3 being phylogenetically different to type 1 and 2. The same
CC degree of differentiation was noted in the NS-5 (see AAR37923-26) and core
CC region between type 3 and type 1 sequences.
CC (Updated on 25-MAR-2003 to correct PN field.)

XX SQ Sequence 117 AA;

Query Match 83.4%; Score 583; DB 14; Length 117;

Best Local Similarity 95.7%; Pred. No. 4.3e-60;

Matches 112; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 8 HPITKTYIMACMSADLEVTSTWVLLGCVLAALAAAYCLSVGCVVIVGHIELGKRPALVPDK 67

DB 1 HPIAKVLMACMSADLEVTSTWVLLGCVLAALAAAYCLSVGCVVIVGHIELGKRPALVPDK 60

QY 68 EVLYQQYDEMECSQAAPYIEQAQVIAHOFKGVGLGLLQRTAQOAVIEPIVTTNQ 124

DB 61 EVLYQQYDEMECSQAAPYIEQAQVIAHOFKGVGLGLLQRTAQOAVIEPIVATNQ 117

RESULT 14

AAW77397
ID AAW77397 standard; Protein; 3011 AA.

XX AC AAW77397;

XX DT 11-JAN-1999 (first entry)

XX DE Hepatitis C virus H77 polyprotein.

XX KW HCV; therapy; diagnosis; vector; gene therapy; vaccine.

XX OS Hepatitis C virus isolate H77.

XX PN WO9839031-A1.

XX PD 11-SEP-1998.

XX PF 26-FEB-1998; 98WO-US04428.

XX PR 04-MAR-1997; 97US-0811566.

XX PA (UNIV) UNIV WASHINGTON.

XX PI Kolykhalov AA, Rice CM;

XX DR WPI; 1998-520770/44.

XX DR N-PSDB; AAV59361.

XX New hepatitis C virus nucleic acid clones - comprising a 5'-terminal
PT conserved sequence, an open reading frame encoding functional
PT components and a 3'-terminal conserved sequence

XX PS Disclosure; Page 104-115; 209pp; English.

XX This is the amino acid sequence of the polyprotein encoded by
CC hepatitis C virus (HCV) isolate H77 (see AAV59361). Its cleavage
CC products form functional components of HCV virus particles and
CC RNA replication machinery. A genetically engineered HCV nucleic
CC acid clone is claimed that comprises at least a functional
CC portion of the HCV H77 nucleic acid sequence. The invention relates

CC to the determination of functional HCV genomic RNA sequences, to
CC construction of infectious HCV DNA clones, and to the use of the
CC clones, or their derivatives, in therapeutic, vaccine and
CC diagnostic applications. The invention is also directed to HCV
CC vectors, e.g. for gene therapy or gene vaccines. The products and
CC methods can also be used for identifying cell lines or animals that
CC are permissive for infection with HCV, for studying HCV infection,
CC isolating functional components of HCV, and for screening for
CC agents capable of modulating HCV replication in vitro and in vivo.

XX SQ Sequence 3011 AA;

Query Match 77.1%; Score 539; DB 19; Length 3011;

Best Local Similarity 75.9%; Pred. No. 4.7e-53;

Matches 101; Conservative 10; Mismatches 22; Indels 0; Gaps 0;

QY 1 ONEICLTHPTIKYIMACMSADLEVTSTWVLLGCVLAALAAAYCLSVGCVVIVGHIELGK 60

DB 1632 QNEVTLTHPTIKYIMTCMSADLEVTSTWVLLGCVLAALAAAYCLSTGCVVIVGRVLSGK 1691

QY 61 PAIVPDKVLYQQYDEMECSQAAPYIEQAQVIAHOFKGVGLGLLQRTAQOAVIEPIVT 120

DB 1692 PAIIPDREVLVYQBFDEMECSQHLPIYIEQGMILAEOFKQKALGLLQRTASHOAEVITPAVQ 1751

QY 121 TNMQKLEAFWHKH 133

DB 1752 TNMQKLEFVFAKH 1764

RESULT 15

ABP71460

ID ABP71460 standard; Protein; 3011 AA.

XX AC ABP71460;

XX DT 15-MAY-2003 (first entry)

XX DE Amino acid sequence of HCV polyprotein.

XX KW HCV; hepatic disorder; envelope glycoprotein; DC-SIGN; DC-SIGNR;

XX KW hepatoprotective; virucide; cytostatic; protein therapy; polyprotein.

XX OS Hepatitis C virus.

XX PN WO2003000024-A2.

XX PD 03-JAN-2003.

XX PF 26-JUN-2002; 2002WO-US20875.

XX PR 26-JUN-2001; 2001US-0891894.

XX PA (PROG-) PROGENICS PHARM INC.

XX PI Olson WC, Maddon PJ;

XX DR WPI; 2003-267852/26.

XX Treating hepatitis, cirrhosis or hepatocellular carcinoma by inhibiting
PT binding between HCV envelope glycoproteins and DC-SIGN/DC-SIGNR
PT proteins on the surface of cells -

XX Claim 100; Fig 3; 165pp; English.

XX The invention relates to methods and agents for diagnosing and treating
CC Hepatitis C Virus (HCV) infections and other hepatic disorders based on
CC binding between HCV envelope glycoproteins and DC-SIGN and DC-SIGNR
CC proteins on the surface of cells. The methods may be used to identify
CC agents (antibodies, peptides and other non-peptidyl agents) that may be
CC administered to treat hepatitis, cirrhosis or hepatocellular carcinomas
CC and to diagnose these disease. The present sequence represents the
CC amino acid sequence of HCV polyprotein (GenBank No. AF009606).

XX

SQ Sequence 3011 AA:

Query Match 77.1%; Score 539; DB 24; Length 3011;

Best Local Similarity 75.9%; Pred. No. 4.7e-53;

Matches 101; Conservative 10; Mismatches 22; Indels 0; Gaps 0;

OY 1 QNEICLTHPTIKYIMACNSADLEVTSTWVLLGGVLAALAAAYCLSVGCVWIVGHIELGGK 60
Db 1632 QNEVTLTHPTIKYIMTCNSADLEVTSTWVLLGGVLAALAAAYCLSVGCVWIVGRIVLSGK 1691

OY 61 PAIVPDKVLYQOYDEMEECQAAPYIEQAQVIAHQPKGVGLGLLQRTQQQAVIEPIVT 120
Db 1692 PAIIPDREVLYQOYDEMEECQHLPLYIEQGMMLAEQFKQKALGLLQRTASQAEVITPAVQ 1751

OY 121 TNWQKLEAFWHKH 133
Db 1752 TNWQKLEVFNAKH 1764

Search completed: August 29, 2003, 11:17:55
Job time : 51 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 11:03:10 ; Search time 21 Seconds
(without alignments)
267.969 Million cell updates/sec

Title: US-09-638-693-36
Perfect score: 699
Sequence: 1 QNEICLTHPIKTYINACMSA.....VIEPIVTTNQKLEAFWKKH 133

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	572	81.8	128	1	US-08-244-116B-17
2	539	77.1	3012	3	US-08-811-566-2
3	539	77.1	3012	4	US-09-034-756-2
4	534	76.4	313	2	US-08-483-695-45
5	534	76.4	313	2	US-07-965-285-45
6	534	76.4	313	2	US-08-487-231-45
7	534	76.4	313	3	US-09-201-912-45
8	534	76.4	382	3	US-08-444-818-68
9	534	76.4	460	3	US-08-444-818-20
10	534	76.4	592	3	US-08-867-611-47
11	534	76.4	594	3	US-08-867-611-48
12	534	76.4	597	3	US-08-867-611-16
13	534	76.4	597	5	PCT-US92-06965A-21
14	534	76.4	599	3	US-08-867-611-18
15	534	76.4	599	5	PCT-US92-06965A-23
16	534	76.4	613	3	US-08-867-611-49
17	534	76.4	739	3	US-08-444-818-148
18	534	76.4	859	3	US-08-444-818-30
19	534	76.4	971	3	US-08-867-611-52
20	534	76.4	973	3	US-08-867-611-53
21	534	76.4	992	3	US-08-867-611-54
22	534	76.4	1021	1	US-07-910-760-12
23	534	76.4	1021	1	US-08-440-519-12
24	534	76.4	1021	4	US-08-440-549-12
25	534	76.4	2261	3	US-08-444-818-66
26	534	76.4	2436	3	US-08-444-818-75
27	534	76.4	2772	3	US-08-444-818-89

28	534	76.4	2894	2	US-08-466-975A-23
29	534	76.4	2894	2	US-08-391-671A-23
30	534	76.4	2894	3	US-08-467-902A-23
31	534	76.4	2894	3	US-09-275-265-23
32	534	76.4	2894	4	US-09-941-611-23
33	534	76.4	2955	2	US-08-443-260-3
34	534	76.4	2955	3	US-08-442-805A-3
35	534	76.4	2955	3	US-08-443-900A-3
36	534	76.4	2955	3	US-08-444-818-124
37	534	76.4	2955	3	US-08-249-843-3
38	534	76.4	2955	3	US-08-444-818-138
39	534	76.4	3011	1	US-08-440-103-36
40	534	76.4	3011	1	US-08-440-542-36
41	534	76.4	3011	1	US-07-910-760-10
42	534	76.4	3011	1	US-08-440-519-10
43	534	76.4	3011	1	US-08-231-368-36
44	534	76.4	3011	1	US-08-440-210-36
45	534	76.4	3011	2	US-08-833-678A-6

ALIGNMENTS

RESULT 1
US-08-244-116B-17
; Sequence 17, Application US/08244116B
; Patent No. 5763159
; GENERAL INFORMATION:
; APPLICANT: Simmonds, Peter
; APPLICANT: Chan, Shiu-Wan
; APPLICANT: Yap, Peng L.
; TITLE OF INVENTION: Hepatitis-C Virus Testing
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell, Seltzer, Park & Gibson, P.A.
; STREET: 1211 East Morehead Street
; CITY: Charlotte
; STATE: No. 5763159th Carolina
; COUNTRY: United States
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0. Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/244,116B
; FILING DATE: 15-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB92/02143
; FILING DATE: 20-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 1749-125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 704-377-1561
; TELEFAX: 704-334-2014
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 128 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: yes
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Hepatitis-C virus
US-08-244-116B-17

Query Match 81.8%; Score 572; DB 1; Length 128;

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
;; ADDRESSEE: Dunner
;; STREET: 1300 I Street, N.W.
;; CITY: Washington
;; STATE: DC
;; COUNTRY: USA
;; ZIP: 20005-3315
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/483,695
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/07/965,285
;; FILING DATE: 18-MAR-1993
;; APPLICATION NUMBER: FR 91 06 882
;; FILING DATE: 06-JUN-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Meyers, Kenneth J.
;; REGISTRATION NUMBER: 25,146
;; REFERENCE/DOCKET NUMBER: 05286-0001-00000
;; TELEPHONE: 202-408-4000
;; TELEFAX: 202-408-4400
;; INFORMATION FOR SEQ ID NO: 45:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 313 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-483-695-45

Query Match 76.4%; Score 534; DB 2; Length 313;
Best Local Similarity 74.4%; Pred. No. 2.7e-54;
Matches 99; Conservative 12; Mismatches 22; Indels 0; Gaps 0;

QY 1 QNEICLTHPTIKYIMACHSADLEVTSTWLVGGVLAALAAAYCLSVGCVVIGHIELGK 60
DB 178 QNEITLTHPTVKYIMTCSADLEVTSTWLVGGVLAALAAAYCLSTGCVVIGRVVLSK 237

QY 61 PAIVDPKEVLVQYDEMEECQAAPYIEQAQVIAHQFKGVGLGLLQRTAQQAQVIEPIVT 120
DB 238 PAIIPDREVLVREFDEMEECQHLPLYIEQGMMLAEQFKOKALGLLQRTASQAQVIEPAVE 297

QY 121 TNWQKLEAFWHKH 133
DB 298 TNWQKLETFWAKH 310

;; ZIP: 20005-3315
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/965,285
;; FILING DATE: 18-MAR-1993
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: FR 91 06 882
;; FILING DATE: 06-JUN-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Meyers, Kenneth J.
;; REGISTRATION NUMBER: 25,146
;; REFERENCE/DOCKET NUMBER: 05286-0001-00000
;; TELEPHONE: 202-408-4000
;; TELEFAX: 202-408-4400
;; INFORMATION FOR SEQ ID NO: 45:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 313 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-07-965-285-45

Query Match 76.4%; Score 534; DB 2; Length 313;
Best Local Similarity 74.4%; Pred. No. 2.7e-54;
Matches 99; Conservative 12; Mismatches 22; Indels 0; Gaps 0;

QY 1 QNEICLTHPTIKYIMACHSADLEVTSTWLVGGVLAALAAAYCLSVGCVVIGHIELGK 60
DB 178 QNEITLTHPTVKYIMTCSADLEVTSTWLVGGVLAALAAAYCLSTGCVVIGRVVLSK 237

QY 61 PAIVDPKEVLVQYDEMEECQAAPYIEQAQVIAHQFKGVGLGLLQRTAQQAQVIEPIVT 120
DB 238 PAIIPDREVLVREFDEMEECQHLPLYIEQGMMLAEQFKOKALGLLQRTASQAQVIEPAVE 297

QY 121 TNWQKLEAFWHKH 133
DB 298 TNWQKLETFWAKH 310

RESULT 6

US-08-487-231-45
; Sequence 45, Application US/08487231
; Patent No. 5919454
; GENERAL INFORMATION:
; APPLICANT: Brechot, Christian
; APPLICANT: Kremsdorf, Dina
; APPLICANT: Porchon, Colette
; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
; TITLE OF INVENTION: Applications
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,231
; FILING DATE: 07-JUNE-1995

RESULT 5

US-07-965-285-45
; Sequence 45, Application US/07965285
; Patent No. 5879904
; GENERAL INFORMATION:
; APPLICANT: Brechot, Christian
; APPLICANT: Kremsdorf, Dina
; APPLICANT: Porchon, Colette
; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
; TITLE OF INVENTION: Applications
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/965,285
FILING DATE: 18-MAR-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 06 882
FILING DATE: 06-JUN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 05286-0001-02000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 313 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-487-231-45

Query Match 76.4%; Score 534; DB 2; Length 313;
Best Local Similarity 74.4%; Pred. No. 2.7e-54;
Matches 99; Conservative 12; Mismatches 22; Indels 0; Gaps 0;
QY 1 QNEICLTHPTKYIMACMSADLEVTSTWVLGGVLAALAAAYCISVGVVIVGHIELGK 60
DB 178 QNEITLTHPTKYIMTCMSADLEVTSTWVLGGVLAALAAAYCISVGVVIVGRVLSGK 237
QY 61 PAIPVDKEVLYQYDEMEECSSQAAPYIEQAOVIAHQFKGVGLLQRTAQOQAVIEPIVT 120
DB 238 PAIPDREVLVREFDEMEECSSQHLPYIEQGMMLAEQFKQKALGLLQTSRQAEVIAPAVE 297
QY 121 TNQKLEAFWVKH 133
DB 298 TNQKLETFWAKH 310

RESULT 7
US-09-201-912-45
Sequence 45, Application US/09201912.
Patent No. 6210962
GENERAL INFORMATION:
APPLICANT: Brechot, Christian
APPLICANT: Kremsdorf, Dina
APPLICANT: Porchon, Colette
TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
TITLE OF INVENTION: Applications
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/201,912
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/965,285
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 05286-0001-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 313 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-201-912-45
Query Match 76.4%; Score 534; DB 3; Length 313;
Best Local Similarity 74.4%; Pred. No. 2.7e-54;
Matches 99; Conservative 12; Mismatches 22; Indels 0; Gaps 0;
QY 1 QNEICLTHPTKYIMACMSADLEVTSTWVLGGVLAALAAAYCISVGVVIVGHIELGK 60
DB 178 QNEITLTHPTKYIMTCMSADLEVTSTWVLGGVLAALAAAYCISVGVVIVGRVLSGK 237
QY 61 PAIPVDKEVLYQYDEMEECSSQAAPYIEQAOVIAHQFKGVGLLQRTAQOQAVIEPIVT 120
DB 238 PAIPDREVLVREFDEMEECSSQHLPYIEQGMMLAEQFKQKALGLLQTSRQAEVIAPAVE 297
QY 121 TNQKLEAFWVKH 133
DB 298 TNQKLETFWAKH 310

RESULT 8
US-08-444-818-68
Sequence 68, Application US/08444818
Patent No. 6150087
GENERAL INFORMATION:
APPLICANT: Chien, David Y.
APPLICANT: Rutter, William J.
TITLE OF INVENTION: NANV Diagnostics and Vaccines
NUMBER OF SEQUENCES: 777
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: CA
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,818
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/403,590
FILING DATE: 14-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Harbin, Alisa A.
REGISTRATION NUMBER: 33,895
REFERENCE/DOCKET NUMBER: 0110.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508)359-3876
TELEFAX: (508)359-3885
INFORMATION FOR SEQ ID NO: 68:
SEQUENCE CHARACTERISTICS:
LENGTH: 382 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-444-818-68

Query Match 76.4%; Score 534; DB 3; Length 382;
Best Local Similarity 74.4%; Pred. No. 3.5e-54;
Matches 99; Conservative 12; Mismatches 22; Indels 0; Gaps 0;

QY 1 QNEICLTHPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSVGCVVIVGHIELGK 60
DB 78 QNEITLTHPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSVGCVVIVGRVLSGK 137
QY 61 PAIVDPKEVLYQOYDEMECSQAAPYIEQAQVIAHQFKVGLGLLQRTAQQAIVIEPIVT 120
DB 138 PAIIPREVLYREFDEMECSQHLPIYIEQGMMLAEQFKQKALGLLQRTASRQAEVIAPAVQ 197
QY 121 TNWQKLEAFWHKH 133
DB 198 TNWQKLETFWAKH 210

RESULT 9

US-08-444-818-20
; Sequence 20, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANOV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 460 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-444-818-20

Query Match 76.4%; Score 534; DB 3; Length 460;
Best Local Similarity 74.4%; Pred. No. 4.5e-54;
Matches 99; Conservative 12; Mismatches 22; Indels 0; Gaps 0;

QY 1 QNEICLTHPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSVGCVVIVGHIELGK 60
DB 162 QNEITLTHPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSVGCVVIVGRVLSGK 221
QY 61 PAIVDPKEVLYQOYDEMECSQAAPYIEQAQVIAHQFKVGLGLLQRTAQQAIVIEPIVT 120
DB 222 PAIIPREVLYREFDEMECSQHLPIYIEQGMMLAEQFKQKALGLLQRTASRQAEVIAPAVQ 281

QY 121 TNWQKLEAFWHKH 133
DB 282 TNWQKLETFWAKH 294

RESULT 10

US-08-867-611-47
; Sequence 47, Application US/08867611
; Patent No. 6172189
; GENERAL INFORMATION:
; APPLICANT: DEVARE, SUSHIL G
; APPLICANT: DESAI, SURESH M
; APPLICANT: CASEY, JAMES M
; APPLICANT: DAILEY, STEPHEN H
; APPLICANT: DAWSON, GEORGE J
; APPLICANT: GUTIERREZ, ROBIN A
; APPLICANT: LESNIEWSKI, RICHARD R
; APPLICANT: STEWART, JAMES L
; APPLICANT: RUPPRECHT, KEVIN R
; TITLE OF INVENTION: HEPATITIS C ASSAY UTILIZING RECOMBINANT
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES
; STREET: ONE ABBOTT PARK ROAD, CHAD377/AP6D2
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,611
; FILING DATE: 02-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,757
; FILING DATE:
; APPLICATION NUMBER: US/08/179,896
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/572,822
; FILING DATE: 24-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/614,069
; FILING DATE: 07-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/748,561
; FILING DATE: 21-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/748,565
; FILING DATE: 21-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/748,566
; FILING DATE: 21-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: POREMSKI, PRISCILLA E
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 4834.US.P6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 592 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-867-611-47

2

APPLICATION NUMBER: US 07/614,069
FILING DATE: 07-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/748,561
FILING DATE: 21-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/748,565
FILING DATE: 21-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/748,566
FILING DATE: 21-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: FOREMSKI, PRISCILLA E
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 4834.US.P6
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 597 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-867-611-16

Query Match 76.4%; Score 534; DB 3; Length 597;
Best Local Similarity 74.4%; Pred. No. 6.4e-54;
Matches 99; Conservative 12; Mismatches 22; Indels 0; Gaps 0;
QY 1 ONEICLTHPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSVGCVVIVGHIELGK 60
DB 288 ONEITLHPVKYIMTCMSADLEVTSTWVLGGVLAALAAAYCLSTGCVVIVGRVVLGK 347
QY 61 PAIVDPKEVLYQOYDEMEECSSQAAPYIEQAQVIAHQFKVGLGLLQRTAQQAQVIEPIVT 120
DB 348 PAIIPDREVLYREFDEMEECSSQHLPIYIEQGMMLAEQFKQKALGLLQRTASRQAQVIAPAVQ 407
QY 121 TNWQKLEAFWHKH 133
DB 408 TNWQKLETFWAKH 420

RESULT 13
PCT-US92-06965A-21
Sequence 21, Application PC/TUS9206965A
GENERAL INFORMATION:
APPLICANT: DEVARE, S.
APPLICANT: DESAI, S.
APPLICANT: DAILEY, S.
TITLE OF INVENTION: HCV SYNTHETIC PEPTIDE FROM NS1 REGION
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES
STREET: ONE ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: ILLINOIS
COUNTRY: U.S.
ZIP: 60065-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/06965A
FILING DATE: 19920821
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: FOREMSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 4834PC.02
TELECOMMUNICATION INFORMATION:

TELEPHONE: 708-937-6365
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 597 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-06965A-21
Query Match 76.4%; Score 534; DB 5; Length 597;
Best Local Similarity 74.4%; Pred. No. 6.4e-54;
Matches 99; Conservative 12; Mismatches 22; Indels 0; Gaps 0;
QY 1 ONEICLTHPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSVGCVVIVGHIELGK 60
DB 288 ONEITLHPVKYIMTCMSADLEVTSTWVLGGVLAALAAAYCLSTGCVVIVGRVVLGK 347
QY 61 PAIVDPKEVLYQOYDEMEECSSQAAPYIEQAQVIAHQFKVGLGLLQRTAQQAQVIEPIVT 120
DB 348 PAIIPDREVLYREFDEMEECSSQHLPIYIEQGMMLAEQFKQKALGLLQRTASRQAQVIAPAVQ 407
QY 121 TNWQKLEAFWHKH 133
DB 408 TNWQKLETFWAKH 420

RESULT 14
US-08-867-611-18
Sequence 18, Application US/08867611
Patent No. 6172189
GENERAL INFORMATION:
APPLICANT: DEVARE, SUSHIL G
APPLICANT: DESAI, SURESH M
APPLICANT: CASEY, JAMES M
APPLICANT: DAILEY, STEPHEN H
APPLICANT: DAWSON, GEORGE J
APPLICANT: GUTIERREZ, ROBIN A
APPLICANT: LESNIEWSKI, RICHARD R
APPLICANT: STEWART, JAMES L
APPLICANT: RUPPRECHT, KEVIN R
TITLE OF INVENTION: HEPATITIS C ASSAY UTILIZING RECOMBINANT
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES
STREET: ONE ABBOTT PARK ROAD, CHAD377/AP6D2
CITY: ABBOTT PARK
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/867,611
FILING DATE: 02-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/646,757
FILING DATE:
APPLICATION NUMBER: US/08/179,896
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/572,822
FILING DATE: 24-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/614,069
FILING DATE: 07-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/748,561

; FILING DATE: 21-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/748,565
; FILING DATE: 21-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/748,566
; FILING DATE: 21-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKI, PRISCILLA E
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 4834.US.P6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 599 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-867-611-18

Query Match 76.4%; Score 534; DB 3; Length 599;
Best Local Similarity 74.4%; Pred. No. 6.4e-54;
Matches 99; Conservative 12; Mismatches 22; Indels 0; Gaps 0;

Qy 1 QNEICLTHPTTKYIMACMSADLEVTTSTWVLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
Db 290 QNEITLTHPTTKYIMTCMSADLEVTTSTWVLGGVLAALAAAYCLSTGCVVIVGRVLSGK 349

Qy 61 PAIVPDKEVLYQYDEMEECSSQAAPYIEQAQVIAHQFKGVGLGLLQRTATQQQAVIEPIVT 120
Db 350 PAIIPDREVLYREFDEMEECSSQHLPLYIEQGMMLAEQFKQKALGLLQRTASRQAEVIAPAVQ 409

Qy 121 TNMOKLEAFWHKH 133
Db 410 TNMOKLETFWAKH 422

RESULT 15
PCT-US92-06965A-23
; Sequence 23, Application PC/TUS9206965A
; GENERAL INFORMATION:
; APPLICANT: DEVARE, S.
; APPLICANT: DESAI, S.
; APPLICANT: DAILEY, S.
; TITLE OF INVENTION: HCV SYNTHETIC PEPTIDE FROM NS1 REGION
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES
; STREET: ONE ABBOTT PARK ROAD
; CITY: ABBOTT PARK
; STATE: ILLINOIS
; COUNTRY: U.S.
; ZIP: 60065-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06965A
; FILING DATE: 19920821
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKI, PRISCILLA E.
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 4834PC.02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 599 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US92-06965A-23

Query Match 76.4%; Score 534; DB 5; Length 599;
Best Local Similarity 74.4%; Pred. No. 6.4e-54;
Matches 99; Conservative 12; Mismatches 22; Indels 0; Gaps 0;

Qy 1 QNEICLTHPTTKYIMACMSADLEVTTSTWVLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
Db 290 QNEITLTHPTTKYIMTCMSADLEVTTSTWVLGGVLAALAAAYCLSTGCVVIVGRVLSGK 349

Qy 61 PAIVPDKEVLYQYDEMEECSSQAAPYIEQAQVIAHQFKGVGLGLLQRTATQQQAVIEPIVT 120
Db 350 PAIIPDREVLYREFDEMEECSSQHLPLYIEQGMMLAEQFKQKALGLLQRTASRQAEVIAPAVQ 409

Qy 121 TNMOKLEAFWHKH 133
Db 410 TNMOKLETFWAKH 422

Search completed: August 29, 2003, 11:18:57
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 11:16:59 ; Search time 184 Seconds
(without alignments)
98.870 Million cell updates/sec

Title: US-09-638-693-36

Perfect score: 699

Sequence: 1 QNEICLTHPTIKYIMACMSA.....VIEPIVTTNQKLEAFWHKH 133

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 510680 seqs, 136781880 residues

Total number of hits satisfying chosen parameters: 510680

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	699	100.0	133	11	US-09-899-046-36
2	699	100.0	133	11	US-09-878-281-36
3	691	98.9	133	11	US-09-899-046-38
4	691	98.9	133	11	US-09-899-046-40
5	691	98.9	133	11	US-09-878-281-38
6	691	98.9	133	11	US-09-878-281-40
7	691	98.9	209	11	US-09-899-046-223
8	691	98.9	209	11	US-09-878-281-223
9	672	96.1	133	11	US-09-899-046-32
10	672	96.1	133	11	US-09-899-046-34
11	672	96.1	133	11	US-09-878-281-32
12	672	96.1	133	11	US-09-878-281-34
13	539	77.1	3011	9	US-09-742-659-4
14	539	77.1	3011	11	US-09-891-894-3
15	539	77.1	3011	12	US-10-184-150-3

Sequence 2, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 23, Appli
Sequence 23, Appli
Sequence 2, Appli
Sequence 20, Appli
Sequence 9, Appli
Sequence 1, Appli
Sequence 20, Appli
Sequence 20, Appli
Sequence 1, Appli
Sequence 20, Appli
Sequence 6, Appli
Sequence 4, Appli
Sequence 11, Appli
Sequence 18, Appli
Sequence 2, Appli
Sequence 2, Appli
Sequence 9, Appli
Sequence 16, Appli
Sequence 42, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 198, App
Sequence 198, App
Sequence 270, App
Sequence 270, App

16 539 77.1 3012 10 US-09-238-076-2
17 539 77.1 3012 11 US-09-995-937-2
18 539 77.1 3012 11 US-09-917-563-2
19 534 76.4 2894 10 US-09-941-611-23
20 534 76.4 2894 15 US-10-044-995-23
21 534 76.4 3011 9 US-09-916-359-2
22 534 76.4 3011 10 US-09-238-076-20
23 534 76.4 3011 10 US-09-952-572-9
24 534 76.4 3011 10 US-09-929-955-1
25 534 76.4 3011 10 US-09-747-419-20
26 534 76.4 3011 11 US-09-995-937-20
27 534 76.4 3011 11 US-09-917-563-20
28 534 76.4 3011 14 US-10-104-966-1
29 534 76.4 3011 15 US-10-259-275-20
30 534 76.4 3011 16 US-10-232-643-6
31 492 70.4 1692 11 US-09-919-901-4
32 492 70.4 1692 11 US-09-919-901-11
33 492 70.4 1692 11 US-09-919-901-18
34 492 70.4 2201 14 US-10-085-476-2
35 492 70.4 2307 11 US-09-919-901-2
36 492 70.4 2307 11 US-09-919-901-9
37 492 70.4 2307 11 US-09-919-901-16
38 490 70.1 1985 15 US-10-259-275-42
39 490 70.1 2985 15 US-10-259-275-40
40 484 69.2 2201 12 US-10-309-561-3
41 484 69.2 2201 14 US-10-029-907-3
42 466 66.7 484 11 US-09-899-046-198
43 466 66.7 484 11 US-09-878-281-198
44 462 66.1 481 11 US-09-899-046-270
45 462 66.1 481 11 US-09-878-281-270

ALIGNMENTS

RESULT 1

US-09-899-046-36
; Sequence 36, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 133 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-899-046-36

Query Match 100.0%; Score 699; DB 11; Length 133;
Best Local Similarity 100.0%; Pred. No. 3.6e-72;
Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QNEICLTHPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSVGVVIGHIELGK 60
Db 1 QNEICLTHPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSVGVVIGHIELGK 60
QY 61 PAIVDPKEVLYOQYDEMECECSQAAPYIEQAQVIAHQFGKVLGLLQRAFOOQAVIEPIVT 120
|||||

Db 61 PAIVDPKEVLYQYDEMEECQAAPYIEQAQVIAHQFKGKVLGLLQRTQQQAVIEPIVT 120
QY 121 TNQKLEAFWKKH 133
Db 121 TNQKLEAFWKKH 133

RESULT 2

US-09-878-281-36
; Sequence 36, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 133 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-878-281-36

Query Match 100.0%; Score 699; DB 11; Length 133;
Best Local Similarity 100.0%; Pred. No. 3.6e-72;
Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 QNEICLTHPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
Db 1 QNEICLTHPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
QY 61 PAIVDPKEVLYQYDEMEECQAAPYIEQAQVIAHQFKGKVLGLLQRTQQQAVIEPIVT 120
Db 61 PAIVDPKEVLYQYDEMEECQAAPYIEQAQVIAHQFKGKVLGLLQRTQQQAVIEPIVT 120
QY 121 TNQKLEAFWKKH 133
Db 121 TNQKLEAFWKKH 133

RESULT 3

US-09-899-046-38
; Sequence 38, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 38:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 133 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-899-046-38

Query Match 98.9%; Score 691; DB 11; Length 133;
Best Local Similarity 99.2%; Pred. No. 3e-71;
Matches 132; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 QNEICLTHPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
Db 1 QNEICLTHPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
QY 61 PAIVDPKEVLYQYDEMEECQAAPYIEQAQVIAHQFKGKVLGLLQRTQQQAVIEPIVT 120
Db 61 PAIVDPKEVLYQYDEMEECQAAPYIEQAQVIAHQFKGKVLGLLQRTQQQAVIEPIVT 120
QY 121 TNQKLEAFWKKH 133
Db 121 TNQKLEAFWKKH 133

RESULT 4

US-09-899-046-40
; Sequence 40, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 133 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-899-046-40

Query Match 98.9%; Score 691; DB 11; Length 133;
Best Local Similarity 99.2%; Pred. No. 3e-71;
Matches 132; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 QNEICLTHPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
Db 1 QNEICLTHPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
QY 61 PAIVDPKEVLYQYDEMEECQAAPYIEQAQVIAHQFKGKVLGLLQRTQQQAVIEPIVT 120
Db 61 PAIVDPKEVLYQYDEMEECQAAPYIEQAQVIAHQFKGKVLGLLQRTQQQAVIEPIVT 120
QY 121 TNQKLEAFWKKH 133
Db 121 TNQKLEAFWKKH 133

RESULT 5

US-09-878-281-38
; Sequence 38, Application US/09878281
; Publication No. US20030032005A1


```
;
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION NUMBER: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 133 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-878-281-38

Query Match 98.9%; Score 691; DB 11; Length 133;
Best Local Similarity 99.2%; Pred. No. 3e-71;
Matches 132; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QNEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSGCVVIVGHIELGGK 60
Db 1 QNEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSGCVVIVGHIELGGK 60
QY 61 PAIVDPKEVLYQYDEMECSQAAPYIEQAQVIAHOFKGVLLGLLQRTQQQAVIEPIVT 120
Db 61 PAIVDPKEVLYQYDEMECSQAAPYIEQAQVIAHOFKGVLLGLLQRTQQQAVIEPIVT 120
QY 121 TNWQKLEAFWPKH 133
Db 121 TNWQKLEAFWPKH 133

RESULT 6
US-09-878-281-40
; Sequence 40, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION NUMBER: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 133 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-878-281-40

Query Match 98.9%; Score 691; DB 11; Length 133;
Best Local Similarity 99.2%; Pred. No. 3e-71;
Matches 132; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QNEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSGCVVIVGHIELGGK 60
Db 1 QNEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSGCVVIVGHIELGGK 60
QY 61 PAIVDPKEVLYQYDEMECSQAAPYIEQAQVIAHOFKGVLLGLLQRTQQQAVIEPIVT 120
Db 61 PAIVDPKEVLYQYDEMECSQAAPYIEQAQVIAHOFKGVLLGLLQRTQQQAVIEPIVT 120
QY 121 TNWQKLEAFWPKH 133
Db 121 TNWQKLEAFWPKH 133

RESULT 7
US-09-899-046-223
; Sequence 223, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 223:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 209 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-899-046-223

Query Match 98.9%; Score 691; DB 11; Length 209;
Best Local Similarity 99.2%; Pred. No. 5.3e-71;
Matches 132; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QNEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSGCVVIVGHIELGGK 60
Db 77 QNEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSGCVVIVGHIELGGK 136
QY 61 PAIVDPKEVLYQYDEMECSQAAPYIEQAQVIAHOFKGVLLGLLQRTQQQAVIEPIVT 120
Db 137 PAIVDPKEVLYQYDEMECSQAAPYIEQAQVIAHOFKGVLLGLLQRTQQQAVIEPIVT 196
QY 121 TNWQKLEAFWPKH 133
Db 197 TNWQKLEAFWPKH 209

RESULT 8
US-09-878-281-223
; Sequence 223, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
```

```
;
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION NUMBER: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 133 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-878-281-38

Query Match 98.9%; Score 691; DB 11; Length 133;
Best Local Similarity 99.2%; Pred. No. 3e-71;
Matches 132; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QNEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSGCVVIVGHIELGGK 60
Db 1 QNEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSGCVVIVGHIELGGK 60
QY 61 PAIVDPKEVLYQYDEMECSQAAPYIEQAQVIAHOFKGVLLGLLQRTQQQAVIEPIVT 120
Db 61 PAIVDPKEVLYQYDEMECSQAAPYIEQAQVIAHOFKGVLLGLLQRTQQQAVIEPIVT 120
QY 121 TNWQKLEAFWPKH 133
Db 121 TNWQKLEAFWPKH 133

RESULT 7
US-09-899-046-223
; Sequence 223, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 223:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 209 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-899-046-223

Query Match 98.9%; Score 691; DB 11; Length 209;
Best Local Similarity 99.2%; Pred. No. 5.3e-71;
Matches 132; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QNEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSGCVVIVGHIELGGK 60
Db 77 QNEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSGCVVIVGHIELGGK 136
QY 61 PAIVDPKEVLYQYDEMECSQAAPYIEQAQVIAHOFKGVLLGLLQRTQQQAVIEPIVT 120
Db 137 PAIVDPKEVLYQYDEMECSQAAPYIEQAQVIAHOFKGVLLGLLQRTQQQAVIEPIVT 196
QY 121 TNWQKLEAFWPKH 133
Db 197 TNWQKLEAFWPKH 209

RESULT 8
US-09-878-281-223
; Sequence 223, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 223:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 209 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-878-281-223

Query Match          98.1%; Score 691; DB 11; Length 209;
Best Local Similarity 93.2%; Pred. No. 5.3e-71;
Matches 132; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ONEICLTHPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
DB 77 ONEICLTHPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 136
QY 61 PAIVDPKEVLYQQYDEMECSQAAPYIEQAQVIAHQFKVGLGLLQORATQQQAVIEPIVT 120
DB 137 PAIVDPKEVLYQQYDEMECSQAAPYIEQAQVIAHQFKVGLGLLQORATQQQAVIEPIVT 196
QY 121 TNNOKLEAFWKKH 133
DB 197 TNNOKLEAFWKKH 209

RESULT 9
US-09-899-046-32
; Sequence 32, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 133 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-899-046-34

Query Match          96.1%; Score 672; DB 11; Length 133;
Best Local Similarity 93.2%; Pred. No. 4.4e-69;
Matches 124; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 ONEICLTHPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
DB 1 ONEICLTHPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
QY 61 PAIVDPKEVLYQQYDEMECSQAAPYIEQAQVIAHQFKVGLGLLQORATQQQAVIEPIVT 120
DB 61 PALVPDKEVLYQQYDEMECSQAAPYIEQAQVIAHQFKVGLGLLQORATQQQAVIEPIVT 120
QY 121 TNNOKLEAFWKKH 133
DB 121 SNQKLETFWKKH 133

RESULT 11
US-09-878-281-32
; Sequence 32, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/878,281
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 133 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```


; Publication No. US20030134297A1
; GENERAL INFORMATION:
; APPLICANT: Olson, William
; APPLICANT: Maddon, Paul
; TITLE OF INVENTION: USES OF DC-SIGN AND DC-SIGNR FOR INHIBITING HEPATITIS C VIRUS INFECTION
; FILE REFERENCE: 2048/64896-A/JPW/MAF/DJK
; CURRENT APPLICATION NUMBER: US/10/184,150
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: 09/891,894
; PRIOR FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 3.
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 3011
; TYPE: PRT
; ORGANISM: Hepatitis C virus
US-10-184-150-3

Query Match 77.1%; Score 539; DB 12; Length 3011;
Best Local Similarity 75.9%; Pred. No. 4.2e-52;
Matches 101; Conservative 10; Mismatches 22; Indels 0; Gaps 0;
QY 1 QNEICLTHPITKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
DB 1632 QNEVTLTHPITKYIMTCMSADLEVTSTWVLGGVLAALAAAYCLSVGCVVIVGRVLSGK 1691
QY 61 PAIYDPREVLYQOYDEMEECSSQAAPYTEQAQVIAHOFKGVGLGLLQATQQQAVIEPIVT 120
DB 1692 PAIIPDREVLYQEFDEMEECSSQHLPLYEQGMMLAEQFKQKALGLLQATASQREVIITPAVQ 1751
QY 121 TNWQKLEAFWHKH 133
DB 1752 TNWQKLEVFVWAKH 1764

Search completed: August 29, 2003, 11:22:48
Job time : 185 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 29, 2003, 11:24:25 ; Search time 3075 Seconds

(without alignments)
1769.425 Million cell updates/sec

Title: US-09-638-693-36

Perfect score: 699

Sequence: 1 QNEICLTHPTIKYIMACMSA.....VIEPIVTINWQKLEAFWHKH 133

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV-xlh

-MODEL=frame_p2n.model -US09638693/runat_25082003_173259_21649/app.query.fasta_1.327

-Q/cgn2_1/USPTO.spool/US09638693/runat_25082003_173259_21649/app.query.fasta_1.327

-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0

-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45

-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL

-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USER=US09638693@cgn_1_1_3508@runat_25082003_173259_21649 -NCPU=6 -ICPU=3

-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG

-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.or.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htg_mus.*
34: em.htg_pln.*
35: em.htg_rod.*
36: em.htg_mam.*
37: em.htg_vrt.*
38: em_sy.*
39: em_hngo_hum.*
40: em_hngo_mus.*
41: em_hngo_other.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	699	100.0	401	6	A40635 Sequence 35
2	699	100.0	401	6	AX031613 Sequence
3	699	100.0	401	6	AX031883 Sequence
4	699	100.0	401	6	AX032153 Sequence
5	699	100.0	401	6	BD172141 New sequ
6	691	98.9	401	6	A40637 Sequence 37
7	691	98.9	401	6	A40639 Sequence 39
8	691	98.9	401	6	AX031615 Sequence
9	691	98.9	401	6	AX031617 Sequence
10	691	98.9	401	6	AX031885 Sequence
11	691	98.9	401	6	AX031887 Sequence
12	691	98.9	401	6	AX032155 Sequence
13	691	98.9	401	6	AX032157 Sequence
14	691	98.9	401	6	BD172142 New sequ
15	691	98.9	401	6	BD172143 New sequ
16	691	98.9	401	6	A40822 Sequence 22
17	691	98.9	401	6	AX031800 Sequence
18	691	98.9	401	6	AX032070 Sequence
19	691	98.9	401	6	AX032340 Sequence
20	691	98.9	401	6	BD172226 New sequ
21	691	98.9	401	6	HPCNS4EE
22	675	96.6	9425	14	AF046866 Hepatitis C
23	674	96.4	1000	14	HCVNS34
24	674	96.4	9390	14	HCVCENS1
25	673	96.3	9456	14	HPCEGS
26	672	96.1	401	6	A40631
27	672	96.1	401	6	A40633 Sequence 31
28	672	96.1	401	6	AX031609 Sequence 33
29	672	96.1	401	6	AX031611 Sequence
30	672	96.1	401	6	AX031879 Sequence
31	672	96.1	401	6	AX031881 Sequence
32	672	96.1	401	6	AX032149 Sequence
33	672	96.1	401	6	AX032151 Sequence
34	672	96.1	401	6	BD172139 New sequ
35	672	96.1	401	6	BD172140 New sequ
36	672	96.1	401	14	HPCNS4G
37	666	95.3	414	14	HCUL4269
38	661	94.6	9454	14	HPCCK3A
39	654	93.6	414	14	HCUL4275
40	654	93.6	414	14	HCUL4277
41	616	88.1	580	14	AF042791
42	605	86.6	9444	6	E10839
43	605	86.6	9444	6	E10840
44	605	86.6	9444	6	E10841
45	605	86.6	9444	14	HPCFCG

ALIGNMENTS

RESULT 1

A40635
LOCUS A40635 401 bp DNA linear PAT 05-MAR-1997
DEFINITION Sequence 35 from Patent WO9425601.
ACCESSION A40635
VERSION A40635.1 GI:2296670
KEYWORDS
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 401)
AUTHORS Maertens,G. and Stuyver,L.
TITLE NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE AS
JOURNAL THERAPEUTIC AND DIAGNOSTIC AGENTS
COMMENT Patent: WO 9425601-A 35 10-NOV-1994;
INNOGENETICS NV (BE)
Other publication CA 2139100 941110
Other publication AU 6722294 941121
Other publication CN 1108030 950906
Other publication FI 946066 941223
Other publication NO 944967 941221
Other publication JP 7508423T 950921.
FEATURES
source Location/Qualifiers
1..401
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/clone="BR36-20-164"
3..401
/note="unnamed protein product; protein sequence is in
conflict with the conceptual translation"
/codon_start=1
/protein_id="CAA02504.1"
/db_xref="GI:2296671"
/translation="ONEICLTHTPTTKYIMACMSADLEVTSTWVLGGVLAALAAAYCL
SVCGVTVGHIELGKPAIVDPKVELYQYDEMECSQAAPYIEQAQVIAHQFQKGV
LLQRATQQAQVIEPIVTTNWKLEAFWFKH"
BASE COUNT 110 a 100 c 101 g 90 t
ORIGIN
Alignment Scores:
Pred. No.: 4.39e-76 Length: 401
Score: 699.00 Matches: 133
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-638-693-36 (1-133) x A40635 (1-401)
Qy 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
Db 3 CAAAATGAAATCTGCTTGACACACCCCATCACAATAATACATCATGCGCATGTCAGCT 62
Qy 21 AspleuGluValThrSerThrTrpValLeuLeuGlyValLeuAlaLeuAla 40
Db 63 GATCTGGAAGTAACACACACACCTGGGTTTTCCTTGAGGGGTCTCTCGCGGGCCCTAGCG 122
Qy 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyGlyLys 60
Db 123 GCCTACTGCTTGCAGTCGGTTGTGTGATGTTGGTGCATATCGAGCTGGGGGCGAAG 182
Qy 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGluCys 80
Db 183 CCGGCAATCTTCAGACAAAGAGGTGTTGTATCAACAATACATGAGATGGAAGAGTGC 242
Qy 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100
Db 243 TCACAAGCTGCCCCATATATCGAACAAAGCTCAGGTAATAGCTCACCAGTTCAAGGGGAAA 302
Qy 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValThr 120
Db 303 GTCTTGGATTGTCGAGCGAGCCACCAACAAAGCTGTCTATTGAGCCCATAGTAAC 362

Qy 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
Db 363 ACCAACTGGCAAAAGCTTGAGCCCTTTTGGCACAGCAT 401
RESULT 2
AX031613 401 bp DNA linear PAT 20-SEP-2000
LOCUS AX031613
DEFINITION Sequence 35 from Patent EP1004670.
ACCESSION AX031613
VERSION AX031613.1 GI:10278850
KEYWORDS
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1
AUTHORS Maertens,G. and Stuyver,L.
TITLE Sequences of hepatitis c virus genotypes and their use as
JOURNAL therapeutic and diagnostic agents
COMMENT Patent: EP 1004670-A 35 31-MAY-2000;
INNOGENETICS NV (BE)
FEATURES
source Location/Qualifiers
1..401
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/clone="BR36-20-164"
3..401
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC09682.1"
/db_xref="GI:10278851"
/translation="ONEICLTHTPTTKYIMACMSADLEVTSTWVLGGVLAALAAAYCL
SVCGVTVGHIELGKPAIVDPKVELYQYDEMECSQAAPYIEQAQVIAHQFQKGV
LLQRATQQAQVIEPIVTTNWKLEAFWFKH"
BASE COUNT 110 a 100 c 101 g 90 t
ORIGIN
Alignment Scores:
Pred. No.: 4.39e-76 Length: 401
Score: 699.00 Matches: 133
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-638-693-36 (1-133) x AX031613 (1-401)
Qy 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
Db 3 CAAAATGAAATCTGCTTGACACACCCCATCACAATAATACATCATGCGCATGTCAGCT 62
Qy 21 AspleuGluValThrSerThrTrpValLeuLeuGlyValLeuAlaLeuAla 40
Db 63 GATCTGGAAGTAACACACACACCTGGGTTTTCCTTGAGGGGTCTCTCGCGGGCCCTAGCG 122
Qy 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyGlyLys 60
Db 123 GCCTACTGCTTGCAGTCGGTTGTGTGATGTTGGTGCATATCGAGCTGGGGGCGAAG 182
Qy 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGluCys 80
Db 183 CCGGCAATCTTCAGACAAAGAGGTGTTGTATCAACAATACATGAGATGGAAGAGTGC 242
Qy 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100
Db 243 TCACAAGCTGCCCCATATATCGAACAAAGCTCAGGTAATAGCTCACCAGTTCAAGGGGAAA 302
Qy 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValThr 120
Db 303 GTCTTGGATTGTCGAGCGAGCCACCAACAAAGCTGTCTATTGAGCCCATAGTAAC 362
121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133

```
Db 363 ACCAACTGGCAAAAGCTTGAGGCCTTTTGGCACAAGCAT 401
RESULT 3
AX031883
LOCUS AX031883 401 bp DNA linear PAT 20-SEP-2000
DEFINITION Sequence 35 from Patent EP0984068.
ACCESSION AX031883
VERSION AX031883.1 GI:10279033
KEYWORDS
SOURCE unidentified
ORGANISM unclassified.
REFERENCE
1 Maertens,G. and Stuyver,L.
AUTHORS Sequences of hepatitis c virus genotypes and their use as
TITLE therapeutic and diagnostic agents
JOURNAL Patent: EP 0984068-A 35 08-MAR-2000;
INNOGENETICS NV (BE)
FEATURES
source
1. .401
Location/Qualifiers
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/clone="BR36-20-164"
<3. .>401
CDS
1. .401
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC09759.1"
/db_xref="GI:10279034"
/translation="QNEICLTHPIITKYIMACMSADLEVTTSWVLLGGVLAALAAAYCL
SYGCVIVIGHIELGKPAIVDPKELYQYDEMECSQAAPYIEQAQVIAHQFKGVL
GILQRAQQQAVIEPIVTNWOKLEAFVHKH"
BASE COUNT 110 a 100 c 101 g 90 t
ORIGIN
Alignment Scores:
Pred. No.: 4.39e-76 Length: 401
Score: 699.00 Matches: 133
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-638-693-36 (1-133) x AX031883 (1-401)
QY 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
Db 3 CAAAATGAAATCTGTTGACACACCCCATCACAAAATACATCATGCGATGTCAGCT 62
QY 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLeuAla 40
Db 63 GATCTGGAAGTAACCAACAGCACCTGGGTTTGTGGAGGGTCTCTCGCGCCCTAGCG 122
QY 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyLys 60
Db 123 GCCTACTGCTTCAGTCAGTCGGTGTGTGTGATTGTGGTGCATATCGAGCTGGGGCAAG 182
QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGluCys 80
Db 183 CCGGCAATGCTTCCAGACAAAGAGGTGTGTATCAACAATACATGAGATGGAGAGTGC 242
QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPhelysGlyLys 100
Db 243 TCACAGCTGCCCATATATATCAACAAGCTCAGGTAAATAGCTCACCAGTTCAAGGAAAA 302
QY 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValThr 120
Db 303 GTCTCTGGATTGCTCAGCGAGCCACCCCAACAACAGCTGTCTATTGAGCCCACTAAGT 362
QY 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
Db 363 ACCAACTGGCAAAAGCTTGAGGCCTTTTGGCACAAGCAT 401
RESULT 5
BD172141
Db 363 ACCAACTGGCAAAAGCTTGAGGCCTTTTGGCACAAGCAT 401
RESULT 4
AX032153
LOCUS AX032153 401 bp DNA linear PAT 20-SEP-2000
DEFINITION Sequence 35 from Patent EP0984067.
ACCESSION AX032153
VERSION AX032153.1 GI:10279216
KEYWORDS
SOURCE unidentified
ORGANISM unclassified.
REFERENCE
1 Maertens,G. and Stuyver,L.
AUTHORS Sequences of hepatitis c virus genotypes and their use as
TITLE therapeutic and diagnostic agents
JOURNAL Patent: EP 0984067-A 35 08-MAR-2000;
INNOGENETICS NV (BE)
FEATURES
source
1. .401
Location/Qualifiers
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/clone="BR36-20-164"
<3. .>401
CDS
1. .401
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC09835.1"
/db_xref="GI:10279217"
/translation="QNEICLTHPIITKYIMACMSADLEVTTSWVLLGGVLAALAAAYCL
SYGCVIVIGHIELGKPAIVDPKELYQYDEMECSQAAPYIEQAQVIAHQFKGVL
GILQRAQQQAVIEPIVTNWOKLEAFVHKH"
BASE COUNT 110 a 100 c 101 g 90 t
ORIGIN
Alignment Scores:
Pred. No.: 4.39e-76 Length: 401
Score: 699.00 Matches: 133
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-638-693-36 (1-133) x AX032153 (1-401)
QY 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
Db 3 CAAAATGAAATCTGTTGACACACCCCATCACAAAATACATCATGCGATGTCAGCT 62
QY 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLeuAla 40
Db 63 GATCTGGAAGTAACCAACAGCACCTGGGTTTGTGGAGGGTCTCTCGCGCCCTAGCG 122
QY 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyLys 60
Db 123 GCCTACTGCTTCAGTCAGTCGGTGTGTGTGATTGTGGTGCATATCGAGCTGGGGCAAG 182
QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGluCys 80
Db 183 CCGGCAATGCTTCCAGACAAAGAGGTGTGTATCAACAATACATGAGATGGAGAGTGC 242
QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPhelysGlyLys 100
Db 243 TCACAGCTGCCCATATATATCAACAAGCTCAGGTAAATAGCTCACCAGTTCAAGGAAAA 302
QY 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValThr 120
Db 303 GTCTCTGGATTGCTCAGCGAGCCACCCCAACAACAGCTGTCTATTGAGCCCACTAAGT 362
QY 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
Db 363 ACCAACTGGCAAAAGCTTGAGGCCTTTTGGCACAAGCAT 401
```

LOCUS	BD172141	401 bp	DNA	linear	PAT 18-FEB-2003							
DEFINITION	New sequences of hepatitis C virus genotypes for diagnosis, phyloglaxis and therapy.											
ACCESSION	BD172141											
VERSION	BD172141.1	GI:284113439										
KEYWORDS	JP 2002233389-A/18.											
SOURCE	unidentified											
ORGANISM	unclassified.											
REFERENCE	1 (bases 1 to 401)											
AUTHORS	Maertens, G. and Stuyver, L.											
TITLE	New sequences of hepatitis C virus genotypes for diagnosis, phyloglaxis and therapy											
JOURNAL	Patent: JP 2002233389-A 18 20-AUG-2002;											
COMMENT	NV INNOGENETICS SA OS Unidentified PN JP 2002233389-A/18 PD 20-AUG-2002 PF 21-NOV-2001 JP 2001356707 PR 27-APR-1993 EP 93401099.2, 05-AUG-1993 EP 93402019, 9 PI PC C12N15/09, A61K38/76, A61K39/00, A61K39/395, A61K39/395, A61K48/00. PC A61P31/20, C07K14/18, C07K16/10, C12Q1/68, G01N33/53, G01N33/53, PC G01N33/366, PC G01N33/576, C12N15/00, A61K37/02 CC Strandedness: Single; CC Topology: Linear; CC New sequences of hepatitis C virus genotypes for diagnosis, phyloglaxis and therapy CC CC FH Key FT CDS											
FEATURES	Location/Qualifiers											
Source	1. .401											
	/organism="unidentified"											
	/mol_type="genomic DNA"											
	/db_xref="taxon:32644"											
BASE COUNT	110 a 100 c 101 g 90 t											
ORIGIN												
Alignment Scores:												
Pred. No.:	4.39e-76	Length:	401									
Score:	699.00	Matches:	133									
Percent Similarity:	100.00%	Conservative:	0									
Best Local Similarity:	100.00%	Mismatches:	0									
Query Match:	100.00%	Indels:	0									
DB:	6	Gaps:	0									
US-09-638-693-36 (1-133) x BD172141 (1-401)												
QY	1	GlnAsnGluIleCysLeuThrHisProIleThrIleTyrIleMetAlaCysMetSerAla	20									
DB	3	CAAAATGAATCTGCTTGACACACCCCATCACAAATACATCATGCGATGTCAGT	62									
QY	21	AspLeuGluValThrSerThrTrpValLeuLeuGlyValLeuAlaLeuAla	40									
DB	63	GATCTGGAAGTAACACACGACCTGGTGTGCTTGGAGGGTCTCGGGCCCTAGCG	122									
QY	41	AlaTyrCysLeuSerValGlyCysValIleValGlyHisIleGluLeuGlyGlyLys	60									
DB	123	GCCTACTGCTTGTGAGTCGGTGTGTGTGATTGTGGTTCATATCGAGCTGGGGGCAAG	182									
QY	61	ProAlaIleValProAspLysGluValLeuTyrGlnClnPyrAspGluMetGluGlyCys	80									
DB	193	CCGCAATCGTTCAGACAAAGAGGTGTGTATCAACAATACCATGAGATGGAAGAGTGC	242									
QY	81	SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys	100									
DB	243	TCACAAGTGCCTCATATACACAGCTCAGGTAACTAGCTCACCAGTTCAAGGAAAA	302									
QY	101	ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValThr	120									
LOCUS	BD172141	401 bp	DNA	linear	PAT 18-FEB-2003							
DEFINITION	New sequences of hepatitis C virus genotypes for diagnosis, phyloglaxis and therapy.											
ACCESSION	BD172141											
VERSION	BD172141.1	GI:284113439										
KEYWORDS	JP 2002233389-A/18.											
SOURCE	unidentified											
ORGANISM	unclassified.											
REFERENCE	1 (bases 1 to 401)											
AUTHORS	Maertens, G. and Stuyver, L.											
TITLE	New sequences of hepatitis C virus genotypes for diagnosis, phyloglaxis and therapy											
JOURNAL	Patent: JP 2002233389-A 18 20-AUG-2002;											
COMMENT	NV INNOGENETICS SA OS Unidentified PN JP 2002233389-A/18 PD 20-AUG-2002 PF 21-NOV-2001 JP 2001356707 PR 27-APR-1993 EP 93401099.2, 05-AUG-1993 EP 93402019, 9 PI PC C12N15/09, A61K38/76, A61K39/00, A61K39/395, A61K39/395, A61K48/00. PC A61P31/20, C07K14/18, C07K16/10, C12Q1/68, G01N33/53, G01N33/53, PC G01N33/366, PC G01N33/576, C12N15/00, A61K37/02 CC Strandedness: Single; CC Topology: Linear; CC New sequences of hepatitis C virus genotypes for diagnosis, phyloglaxis and therapy CC CC FH Key FT CDS											
FEATURES												


```
Db 243 TCACAGAGTCCCATATATCGAACAAAGCTCAGGTGATAGCTCACCAGTTCAGGAAAAA 302
|||
Qy 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValThr 120
|||
Db 303 GTCCCTGGATTGCTCGAGCGAGCCACCACCAACAAAGCTGCTATTGAGCCCATAGTA 362
|||
Qy 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
|||
Db 363 ACCAACTGGCAAAAGCTTGAGGCCCTTTGGCACAGCAT 401
|||
RESULT 7
A40639
LOCUS
DEFINITION
Sequence 39 from Patent WO9425601.
ACCESSION
A40639
VERSION
A40639.1 GI:2296674
KEYWORDS
unidentified
SOURCE
unidentified
ORGANISM
unclassified.
REFERENCE
1 (bases 1 to 401)
AUTHORS
Maertens, G. and Stuyver, L.
TITLE
NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE AS
THERAPEUTIC AND DIAGNOSTIC AGENTS
JOURNAL
Patent: WO 9425601-A 39 10-NOV-1994;
INNOGENETICS NV (BE)
COMMENT
Other publication CA 2139100 941110
Other publication AU 6722294 941121
Other publication CN 1108030 950906
Other publication FI 946066 941223
Other publication NO 944967 941221
Other publication JP 7508423T 950921.
FEATURES
Location/Qualifiers
1..401
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/clone="BR36-20-165"
3..401
/note="unnamed protein product; Protein sequence is in
conflict with the conceptual translation"
/codon_start=1
/protein_id="CAA02506.1"
/db_xref="GI:2296675"
/translation="ONEICLTPIPKYIMACMSADLEVTTSWLLGGVLAALAYCL
SVGCVVIVGHIELGKPAIVDPKEVLYOQYDEMECSQAAPYIEQAQVIAHQFKEKVL
GLQRATQQQAVIEPIVTNNWOKLEAFWKH"
BASE COUNT 111 a 100 c 100 g 90 t
ORIGIN
Alignment Scores:
Pred. No.: 4,15e-75 Length: 401
Score: 691.00 Matches: 132
Percent Similarity: 99.25% Conservative: 0
Best Local Similarity: 99.25% Mismatches: 1
Query Match: 98.86% Indels: 0
DB: Gaps: 0
US-09-638-693-36 (1-133) x A40639 (1-401)
Qy 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTrIleMetAlaCysMetSerAla 20
|||
Db 3 CAAATGAATTCGTTGACACACCCCATCACAATAATCATGCGCATGTCAGCT 62
|||
Qy 21 AspLeuGluValThrThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLeuAla 40
|||
Db 63 GATCTGGAAGTAACCAACAGCACCTGGGTTTTCITGGAGGGTCTCTCCGCGCCCTAGCG 122
|||
Qy 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyGlyLys 60
|||
Db 123 GCCTACTGCTTGCAGTCGGTGTGTGTGATTGTGGGTGCATATCGAGCTGGGGGCAAG 182
|||
```

```
Qy 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluCys 80
|||
Db 183 CCGCAATCGTTCCAGACAAAGAGGTGTTGTATCAACAATACGATGAGATGGAAGAGTCC 242
|||
Qy 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100
|||
Db 243 TCACAGAGTCCCATATATCGAACAAAGCTCAGGTGATAGCTCACCAGTTCAGGAAAAA 302
|||
Qy 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValThr 120
|||
Db 303 GTCCCTGGATTGCTCGAGCGAGCCACCACCAACAAAGCTGCTATTGAGCCCATAGTA 362
|||
Qy 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
|||
Db 363 ACCAACTGGCAAAAGCTTGAGGCCCTTTGGCACAGCAT 401
|||
RESULT 8
AX031615
LOCUS
DEFINITION
Sequence 37 from Patent EP1004670.
ACCESSION
AX031615
VERSION
AX031615.1 GI:10278852
KEYWORDS
unidentified
SOURCE
unidentified
ORGANISM
unclassified.
REFERENCE
1
AUTHORS
Maertens, G. and Stuyver, L.
TITLE
Sequences of hepatitis c virus genotypes and their use as
therapeutic and diagnostic agents
JOURNAL
Patent: EP 1004670-A 37 31-MAY-2000;
INNOGENETICS NV (BE)
FEATURES
Location/Qualifiers
1..401
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/clone="BR36-20-166"
<3..>401
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC09683.1"
/db_xref="GI:10278853"
/translation="ONEICLTPIPKYIMACMSADLEVTTSWLLGGVLAALAYCL
SVGCVVIVGHIELGKPAIVDPKEVLYOQYDEMECSQAAPYIEQAQVIAHQFKEKVL
GLQRATQQQAVIEPIVTNNWOKLEAFWKH"
BASE COUNT 110 a 100 c 101 g 90 t
ORIGIN
Alignment Scores:
Pred. No.: 4,15e-75 Length: 401
Score: 691.00 Matches: 132
Percent Similarity: 99.25% Conservative: 0
Best Local Similarity: 99.25% Mismatches: 1
Query Match: 98.86% Indels: 0
DB: Gaps: 0
US-09-638-693-36 (1-133) x AX031615 (1-401)
Qy 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTrIleMetAlaCysMetSerAla 20
|||
Db 3 CAAATGAATTCGTTGACACACCCCATCACAATAATCATGCGCATGTCAGCT 62
|||
Qy 21 AspLeuGluValThrThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLeuAla 40
|||
Db 63 GATCTGGAAGTAACCAACAGCACCTGGGTTTTCITGGAGGGTCTCTCCGCGCCCTAGCG 122
|||
Qy 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyGlyLys 60
|||
Db 123 GCCTACTGCTTGCAGTCGGTGTGTGTGATTGTGGGTGCATATCGAGCTGGGGGCAAG 182
|||
Qy 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluCys 80
|||
```

```

Db      183 CCGCAATCGTTCACAGAAAGAGGTGTGTATCAACAATACGATGAGTGAAGAGTGC 242
Qy      81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100
Db      243 TCACAGCTGCCCATATATCGAACAGCTCAGGTAGTACCTACCAAGTTCAGGAAAAA 302
Qy      101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValThr 120
Db      303 GTCTTGGATTGCTGCAGCGAGCCACCAACAAGCTGTCTATTGAGCCCATAGTAAC 362
Qy      121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
Db      363 ACCAACTGGCAAAAGCTTGAGGCTTTTGGCACAAGCAT 401

RESULT 9
LOCUS   AX031617
DEFINITION Sequence 39 from Patent EP1004670.
ACCESSION AX031617
VERSION   AX031617.1 GI:10278854
KEYWORDS
SOURCE   unidentified
ORGANISM unidentified
REFERENCE 1
AUTHORS  Maertens,G. and Stuyver,L.
TITLE    Sequences of hepatitis c virus genotypes and their use as
JOURNAL  therapeutic and diagnostic agents
PATENT:  EP 1004670-A 39 31-MAY-2000;
INNOGENETICS NV (BE)
FEATURES             Location/Qualifiers
     source
     1..401
     /organism="unidentified"
     /mol_type="genomic DNA"
     /db_xref="taxon:32644"
     /clone="BR36-20-165"
     <3..>401
     /note="unnamed protein product"
     /codon_start=1
     /protein_id="CAC09684.1"
     /db_xref="GI:10278854"
     /translation="ONEICLTHTPTIKYIMACMSADLEVTTSWVLLGGVLAALAAAYCL
     SVGCVTVGHIELGKPAIVDPDKVELVQYDMEECQAAPIYEQAOVIAHQFKEKVL
     GLQRTAQQAIVIEPIVTNNQKLEAFWHKH"
BASE COUNT  111 a 100 c 100 g 90 t
ORIGIN

Alignment Scores:
Pred. No.:      4.15e-75      Length:      401
Score:          691.00      Matches:    132
Percent Similarity: 99.25%      Conservative: 0
Best Local Similarity: 99.25%      Mismatches: 1
Query Match:    98.86%      Indels:    0
DB:             6          Gaps:      0

US-09-638-693-36 (1-133) x AX031617 (1-401)
Qy      1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
Db      3 CAAATGAATCTGCTTGACACACCCCATCACAAATACATCATGGCATCGATGTCAGCT 62
Qy      21 AspLeuGluValThrThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLeuAla 40
Db      63 GATCTGGAAGTAACCAACACAGCAGCTGGGTTTCTTGAGGGGTCTCTCGCGCCCTAGCG 122
Qy      41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyGlyLys 60
Db      123 GCCTACTGCTGTGACGTCGGTGTGTGTGATTGTGGGTCTATCATCGAGCTGGGGGCAAG 182
Qy      61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGluCys 80
Db      183 CCGCAATCGTTCACAGAAAGAGGTGTCTTATCAACAATACGATGAGTGAAGAGTGC 242

```

```

Qy      81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100
Db      243 TCACAGCTGCCCATATATCGAACAGCTCAGGTAGTACCTACCAAGTTCAGGAAAAA 302
Qy      101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValThr 120
Db      303 GTCTTGGATTGCTGCAGCGAGCCACCAACAAGCTGTCTATTGAGCCCATAGTAAC 362
Qy      121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
Db      363 ACCAACTGGCAAAAGCTTGAGGCTTTTGGCACAAGCAT 401

RESULT 10
LOCUS   AX031885
DEFINITION Sequence 37 from Patent EP0984068.
ACCESSION AX031885
VERSION   AX031885.1 GI:10279035
KEYWORDS
SOURCE   unidentified
ORGANISM unidentified
REFERENCE 1
AUTHORS  Maertens,G. and Stuyver,L.
TITLE    Sequences of hepatitis c virus genotypes and their use as
JOURNAL  therapeutic and diagnostic agents
PATENT:  EP 0984068-A 37 08-MAR-2000;
INNOGENETICS NV (BE)
FEATURES             Location/Qualifiers
     source
     1..401
     /organism="unidentified"
     /mol_type="genomic DNA"
     /db_xref="taxon:32644"
     /clone="BR36-20-166"
     <3..>401
     /note="unnamed protein product"
     /codon_start=1
     /protein_id="CAC09760.1"
     /db_xref="GI:10279036"
     /translation="ONEICLTHTPTIKYIMACMSADLEVTTSWVLLGGVLAALAAAYCL
     SVGCVTVGHIELGKPAIVDPDKVELVQYDMEECQAAPIYEQAOVIAHQFKEKVL
     GLQRTAQQAIVIEPIVTNNQKLEAFWHKH"
BASE COUNT  110 a 100 c 101 g 90 t
ORIGIN

Alignment Scores:
Pred. No.:      4.15e-75      Length:      401
Score:          691.00      Matches:    132
Percent Similarity: 99.25%      Conservative: 0
Best Local Similarity: 99.25%      Mismatches: 1
Query Match:    98.86%      Indels:    0
DB:             6          Gaps:      0

US-09-638-693-36 (1-133) x AX031885 (1-401)
Qy      1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
Db      3 CAAATGAATCTGCTTGACACACCCCATCACAAATACATCATGGCATCGATGTCAGCT 62
Qy      21 AspLeuGluValThrThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLeuAla 40
Db      63 GATCTGGAAGTAACCAACACAGCAGCTGGGTTTCTTGAGGGGTCTCTCGCGCCCTAGCG 122
Qy      41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyGlyLys 60
Db      123 GCCTACTGCTGTGACGTCGGTGTGTGTGATTGTGGGTCTATCATCGAGCTGGGGGCAAG 182
Qy      61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGluCys 80
Db      183 CCGCAATCGTTCACAGAAAGAGGTGTCTTATCAACAATACGATGAGTGAAGAGTGC 242
Qy      81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100

```

Db	243	TCACAAGCTGCCCATATATATCGAACACAGCTCAGCTGATAGCTCACCAGTTCACAGGAAAA	302
Qy	101	ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluPheIleValThr	120
Db	303	GTCCCTTGGATTGCTGCAGCGAGCCACCAACAACAAGTCTCATTTGAGCCCATAGTA	362
Qy	121	ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis	133
Db	363	ACCAACTGGCAAAAGGTTGAGGCCTTTTGGCACAAGCAT	401
RESULT 11			
AX031887			
LOCUS	AX031887	401 bp	DNA
DEFINITION	Sequence 39 from Patent EP0984068.		linear
ACCESSION	AX031887		
VERSION	AX031887.1	GI:10279037	
KEYWORDS			
SOURCE	unidentified		
ORGANISM	unidentified		
REFERENCE	unclassified.		
AUTHORS	Maertens, G. and Stuyver, L.		
TITLE	Sequences of hepatitis c virus genotypes and their use as		
JOURNAL	therapeutic and diagnostic agents		
	Patent: EP 0984068-A 39 08-MAR-2000;		
	INNOGENETICS NV (BE)		
FEATURES	Location/Qualifiers		
source	1..401		
	/organism="unidentified"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:32644"		
	/clone="BR36-20-165"		
	<3..>401		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAC09761.1"		
	/db_xref="GI:10279038"		
	/translation="QNEICLTHTPIKYIMACNSADLEVTSTWVLGGVLAALAAYCL		
	SVGCVIVGHIEUGGPAIVDPKEIVLYQYDEMEECSQAAPYIEQAQVIAHQPKVKVL		
	GLLQRATQQAIVPEIVTNNQKLEAFVKKH"		
BASE COUNT	111 a	100 c	100 g
ORIGIN			90 t
Alignment Scores:			
Pred. No.:	4.15e-75	Length:	401
Score:	691.00	Matches:	132
Percent Similarity:	99.25%	Conservative:	0
Best Local Similarity:	99.25%	Mismatches:	1
Query Match:	98.86%	Indels:	0
DB:	6	Gaps:	0
US-09-638-693-36 (1-133) x AX031887 (1-401)			
Qy	1	GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla	20
Db	3	CAAAATGAATCTGCTTGACACACCCCATCACAAATACATCATGCGATGTCAGCT	62
Qy	21	AspLeuGluValThrThrSerThrTrpValLeuLeuGlyClyValLeuAlaLeuAla	40
Db	63	GATCTGGAAGTACCAACACACACCTGGGTTTTCGTTGGAGGGTCTCTCGGGCCCTAGCG	122
Qy	41	AlaTyrCysLeuSerValGlyCysValIleValGlyHisIleGluLeuGlyGlyLys	60
Db	123	GCCTACTGCTTGTTCAGTCGGTGTGTGTGATTGTGGGTCTATATCGAGCTGGGGGCGAAG	182
Qy	61	ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluCys	80
Db	183	CCGCAATCGTTCAGACAAAGAGGTGTTCTATCAACAATACGATGAGATGGAAGAGTGC	242
Qy	81	SerClnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys	100
Db	243	TCACAAGCTGCCCATATATCGAACACAGCTCAGCTATAGCTCACCAGTTCACAGGAAAA	302

Qy	101	ValLeuGlyLeuLeuClnArgAlaThrGlnGlnGlnAlaValAlleGluProIleValThr	120
Db	303	GTCTTTGGATTGCTGCAGCGAGCCACCCAAACAAGCTGTCTATTGAGCCCATAGTAACT	362
Qy	121	ThrAspTrpGlnLysLeuGluAlaPheTrpHisLysHis	133
Db	363	ACCAACTGGCAAAAGCTTGAGGCCCTTTGGCCACAAGCAT	401
RESULT 12			
AX032155			
LOCUS	AX032155	401 bp	DNA linear PAT 20-SEP-2000
DEFINITION	Sequence 37 from Patent EP0984067.		
ACCESSION	AX032155		
VERSION	AX032155.1	GI:10279218	
KEYWORDS	unidentified		
SOURCE	unidentified		
ORGANISM	unclassified.		
REFERENCE	1		
AUTHORS	Maertens, G. and Stuyver, L.		
TITLE	Sequences of hepatitis C virus genotypes and their use as		
	therapeutic and diagnostic agents		
JOURNAL	Patent: EP 0984067-A 37 08-MAR-2000;		
	INNOGENETICS NV (BE)		
FEATURES	Location/Qualifiers		
source	1..401		
	/organism="unidentified"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:32644"		
	/clone="BR36-20-166"		
	<3..>401		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAC09836.1"		
	/db_xref="GI:10279219"		
	/translation="QNEICLTHPIPTYIMACMSADLEVTSTWVLLGGVLAALAAYCL		
	SVGVVIHIGELGKPAIVDPKELYQYDMEECQAAPYIEQAQVIAHQFKRYL		
	GLLQATQQAQVIEPIVTTNQWLEAFPHKH"		
BASE COUNT	110 a 100 c 101 g 90 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	4.15e-75	Length:	401
Score:	691.00	Matches:	132
Percent Similarity:	99.25%	Conservative:	0
Best Local Similarity:	99.25%	Mismatches:	1
Query Match:	98.86%	Indels:	0
DB:	6	Gaps:	0
US-09-638-693-36 (1-133) x AX032155 (1-401)			
Qy	1	GlnAsnGluIleCysLeuThrHisProIleThrLysTyrrIleMetAlaCysMetSerAla	20
Db	3	CAAAATGAAATCTGCTTGACACACCCCATACAAAAATACATCATGGCATGTGACGT	62
Qy	21	AspLeuGluValThrThrSerThrTrpValLeuLeuGlyGlyValLeuAlaAlaLeuAla	40
Db	63	GATCTGGAAGTAACCAACACGACCTGGGTTTGGCTTGGAGGGGTCTCTCGCGCCCTACGG	122
Qy	41	AlaTyrcysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyGlyLys	60
Db	123	GCCTACTGCTTGTCAGTCGGTGTGTGTGATTGTGGTTCATATCGAGTCGGGGGCAAG	182
Qy	61	ProAlaIleValProAspLysGluValLeuTyrrGlnGlnTyrrAspGluMetGluGluCys	80
Db	183	CCGGCAATCGTCCAGACAAAGAGTGTGTATCAACAATACGATGAGATGGAAGAGTGC	242
Qy	81	SerGlnAlaAlaProTyrrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys	100
Db	243	TCACAAGCTGCCCATATATATCAACACAGCTCAGGTGATAGCTCACCAGTTCAAGGAAAA	302
Qy	101	ValLeuGlyLeuLeuClnArgAlaThrGlnGlnGlnAlaValAlleGluProIleValThr	120

```

Db      303  GTCCCTTGATGTCGACGAGCCACCCCAACAAGCTGTCAATTGAGCCCATAGTAAC 362
Qy      121  ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
Db      363  ACCAACTGGCAAAAGCTTGAGGCCTTTTGGCACAAGCAT 401

RESULT 13
AX032157
LOCUS      AX032157
DEFINITION Sequence 39 from Patent EP0984067.
ACCESSION AX032157
VERSION    AX032157.1 GI:10279220
KEYWORDS
SOURCE     unidentified
ORGANISM   unclassified.
REFERENCE  1
AUTHORS    Maertens,G. and Stuyver,L.
TITLE      Sequences of hepatitis c virus genotypes and their use as
            therapeutic and diagnostic agents
JOURNAL    Patent: EP 0984067-A 39 08-MAR-2000;
            INNOGENETICS NV (BE)
FEATURES   Location/Qualifiers
            source
              1. .401
                /organism="unidentified"
                /mol_type="genomic DNA"
                /db_xref="taxon:32644"
                /clone="BR36-20-165"
              <3. .>401
                /note="unnamed protein product"
                /codon_start=1
                /protein_id="CAC09837.1"
                /db_xref="GI:10279221"
                /translation="ONEICLTHTPTTKVIMACMSADLEVTSTVLLGGVLAALAAAYCL
                SVGCVTVTGHELGKPAIVPDKEVLYQYDEMECSQAAPYIEQAVIAHQFKEKVL
                GLQRATQQQAVIEPIVTNNQKLEAFWRKH"
BASE COUNT 111 a 100 c 100 g 90 t
ORIGIN

Alignment Scores:
Pred. No.: 4.15e-75 Length: 401
Score: 691.00 Matches: 132
Percent Similarity: 99.25% Conservative: 0
Best Local Similarity: 99.25% Mismatches: 1
Query Match: 98.86% Indels: 0
DB: 6 Gaps: 0

US-09-638-693-36 (1-133) x AX032157 (1-401)
Qy      1  GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
Db      3  CAAATGAATCTGCTTGACACACCCCATCACAATAATACATCATGGCATGTCAGCT 62
Qy      21  AspLeuGluValThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLeuAla 40
Db      63  GATCTGGAAGTAACACACAGCACCTGGGTTTGTGTTGGAGGGGTCCTCGCGGCCCTAGCG 122
Qy      41  AlaTyrCysLeuSerValGlyCysValIleValGlyHisIleGluLeuGlyGlyLys 60
Db      123  GCCTACTGCTGTTCAGTCGGTGTGTGTTGATTGTGGTGCATATATCGAGCTGGGGGGCAAG 182
Qy      61  ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGluCys 80
Db      183  CCGGCATCTGCTCCAGACAAGAGGTGTTGTATCAACAAATACGATGATGAGTGAAGAGTGC 242
Qy      81  SerGlnAlaLaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100
Db      243  TCACAAGTGCGCCCATATATCGAACAAAGCTCAGGTAATAGTCTACCAAGTTCAAGGAAAAA 302
Qy      101  ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValThr 120
Db      303  GTCCCTTGATGTCGACGAGCCACCCCAACAAGCTGTCAATTGAGCCCATAGTAAC 362

```

```

Qy      121  ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
Db      363  ACCAACTGGCAAAAGCTTGAGGCCTTTTGGCACAAGCAT 401

RESULT 14
BD172142
LOCUS      BD172142
DEFINITION New sequences of hepatitis C virus genotypes for diagnosis,
            prophylaxis and therapy.
ACCESSION BD172142
VERSION    BD172142.1 GI:28413440
KEYWORDS   JP 2002233389-A/19.
SOURCE     unidentified
ORGANISM   unclassified.
REFERENCE  1 (bases 1 to 401)
AUTHORS    Maertens,G. and Stuyver,L.
TITLE      New sequences of hepatitis C virus genotypes for diagnosis,
            prophylaxis and therapy
JOURNAL    Patent: JP 2002233389-A 19 20-AUG-2002;
            NV INNOGENETICS SA
COMMENT    OS Unidentified
            PN JP 2002233389-A/19
            PD 20-AUG-2002
            PF 21-NOV-2001 JP 2001356707
            PR 27-APR-1993 EP 93401099.2.05-AUG-1993 EP 93402019.9 PI
            GEERT MAERTENS,LIEVEN STUYVER
            PC CL2N15/09.A61K35/76.A61K38/00.A61K39/00.A61K39/395,
            PC A61K48/00,
            PC A61P31/20.C07K14/18.C07K16/10.C12Q1/68.G01N33/53.G01N33/53, PC
            G01N33/566,
            PC G01N33/576.C12N15/00.A61K37/02
            CC Strandedness: Single;
            CC Topology: Linear;
            CC New sequences of hepatitis C virus genotypes for diagnosis,
            CC and therapy
            CC prophylaxis
            CC and therapy
            FT Key
            CDS Location/Qualifiers
              1. .401
                /organism="unidentified"
                /mol_type="genomic DNA"
                /db_xref="taxon:32644"
BASE COUNT 110 a 100 c 101 g 90 t
ORIGIN

Alignment Scores:
Pred. No.: 4.15e-75 Length: 401
Score: 691.00 Matches: 132
Percent Similarity: 99.25% Conservative: 0
Best Local Similarity: 99.25% Mismatches: 1
Query Match: 98.86% Indels: 0
DB: 6 Gaps: 0

US-09-638-693-36 (1-133) x BD172142 (1-401)
Qy      1  GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
Db      3  CAAATGAATCTGCTTGACACACCCCATCACAATAATACATCATGGCATGTCAGCT 62
Qy      21  AspLeuGluValThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLeuAla 40
Db      63  GATCTGGAAGTAACACACAGCACCTGGGTTTGTGTTGGAGGGGTCCTCGCGGCCCTAGCG 122
Qy      41  AlaTyrCysLeuSerValGlyCysValIleValGlyHisIleGluLeuGlyGlyLys 60
Db      123  GCCTACTGCTGTTCAGTCGGTGTGTGTTGATTGTGGTGCATATATCGAGCTGGGGGGCAAG 182
Qy      61  ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGluCys 80
Db      183  CCGGCATCTGCTCCAGACAAGAGGTGTTGTATCAACAAATACGATGATGAGTGAAGAGTGC 242

```

QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPhenylGlyLys 100
|||||
Db 243 TCACAAAGCTGCCCATATATCAACAAGCTCAGGTGATAGCTCACAGTTCAAGGAATA 302
|||||
QY 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValThr 120
|||||
Db 303 GTCCCTGGATTCTCAGCGAGCCACCCCAACAACAGCTGTCTATTGAGCCCATAGTA 362
|||||
QY 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
|||||
Db 363 ACCAACTGGCAAAAGCTTGAGGCCTTTGGCACAGCAT 401
|||||

RESULT 15
BD172143
LOCUS
DEFINITION New sequences of hepatitis C virus genotypes for diagnosis,
prophylaxis and therapy.
BD172143
ACCESSION
VERSION BD172143.1 GI:28413441
KEYWORDS JP 2002233389-A/20.
SOURCE unidentified
ORGANISM unidentified
unclassified.

REFERENCE 1 (bases 1 to 401)
AUTHORS Maertens, G. and Stuyver, L.
TITLE New sequences of hepatitis C virus genotypes for diagnosis,
prophylaxis and therapy
JOURNAL Patent: JP 2002233389-A 20-AUG-2002;
NV INNOGENETICS SA
COMMENT OS Unidentified
PN JP 2002233389-A/20
PD 20-AUG-2002
PF 21-NOV-2001 JP 2001356707
PR 27-APR-1993 EP 93401099.2, 05-AUG-1993 EP 93402019.9 PI
GEERT HAERTENS, LIEVEN STUYVER
PC C12N15/09, A61K35/76, A61K38/00, A61K39/00, A61K39/395, A61K39/395,
A61K48/00,
PC A61P31/20, C07K14/18, C07K16/10, C12Q1/68, G01N33/53, G01N33/53, PC
G01N33/566,
PC G01N33/576, C12N15/00, A61K37/02
CC Strandedness: Single;
CC Topology: Linear;
CC New sequences of hepatitis C virus genotypes for diagnosis,
prophylaxis
CC and therapy
FH Key Location/Qualifiers
FT CDS 1..401
Location/Qualifiers
1..401
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

BASE COUNT 111 a 100 c 100 g 90 t
ORIGIN

Alignment Scores:
Pred. No.: 4 15e-75 Length: 401
Score: 691.00 Matches: 132
Percent Similarity: 99.25% Conservative: 0
Best Local Similarity: 99.25% Mismatches: 1
Query Match: 98.86% Indels: 0
DB: 6 Gaps: 0

US-09-638-693-36 (1-133) x BD172143 (1-401)

QY 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
|||||
Db 3 CAAATGAATCTGCTTGACACACCCCATCACAATAATACATCATGCGATGTCAGCT 62
|||||
QY 21 AspLeuGluValThrThrSerThrTrpValLeuLeuGlyValLeuAlaLeuAla 40
|||||
Db 63 GATCTGGAAGTAACACACACCTGGGTTTGTGGAGGGTCTCTCGCGGCCCTAGCG 122
|||||

QY 41 AlaTyrCysLeuSerValGlyCysValValIleValIleValGlyHisIleGluLeuGlyGlyLys 60
|||||
Db 123 GCCTACTGCTTGTCTAGTCGGTTGTGTGATTGTGGTTCATATCGAGCTGGGGGCAAG 182
|||||
QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGluCys 80
|||||
Db 183 CCGGCAATCGTTCCAGACAAGAGGTTGTGTATCAACAATACGATGAGATGGAAGAGTGC 242
|||||
QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPhenylGlyLys 100
|||||
Db 243 TCACAAAGCTGCCCATATATCAACAAGCTCAGGTANTAGCTCACAGTTCAAGGAATA 302
|||||
QY 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValThr 120
|||||
Db 303 GTCCCTGGATTCTCAGCGAGCCACCCCAACAACAGCTGTCTATTGAGCCCATAGTA 362
|||||
QY 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
|||||
Db 363 ACCAACTGGCAAAAGCTTGAGGCCTTTGGCACAGCAT 401
|||||

Search completed: August 29, 2003, 12:21:36
Job time : 3078 secs

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

```

FT CDS 3..401
FT /*tag= a
FT /product= NS3/NS4 polypeptide.
XX
PN W09425601-A2.
XX
PD 10-NOV-1994.
XX
PF 27-APR-1994; 94WO-EP01323.
XX
PR 27-APR-1993; 93EP-0401099.
XX
PR 05-AUG-1993; 93EP-0402019.
XX
PA (INNO-) INNOGENETICS NV SA.
XX
PI Maertens G, Stuyver L;
XX
DR WPI; 1994-358277/44.
DR P-PSDB; AAR63288.
XX
PT New polynucleotide sequences from hepatitis C virus - and related
PT vectors, polypeptide(s) and antibodies, useful for immunisation,
PT treatment, diagnosis and typing of HCV isolates
XX
PS Claim 2; Page 124-125; 404pp; English.
XX
CC Compositions comprising at least 5, and pref. 8 or more contiguous
CC nucleotides selected from an HCV type 3 genomic sequence, more
CC particularly (i) the region spanning positions 417-957 of the
CC Core/E1 region of HCV subtype 3a; (ii) the region spanning positions
CC 4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
CC positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
CC region spanning positions 8023-8235 of the NS5 region of the BR36
CC subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
CC sequence, may be used as primers to amplify nucleic acid from an
CC isolate belonging to a specific genotype, or as a probe for specific
CC detection/classification of nucleic acid. Polypeptides encoded by
CC the nucleotides in such compositions may be used for immunisation
CC against HCV, for the detection of antibodies directed against HCV
CC and for serotyping. This sequence corresponds to the NS3/NS4
CC region of HCV subtype 3a and is taken from a clone designated
CC BR36-20-164.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 401 BP; 110 A; 100 C; 101 G; 90 T; 0 other;

Alignment Scores:
Pred., No.: 2.74e-78 Length: 401
Score: 699.00 Matches: 133
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-09-638-693-36 (1-133) x AAQ78040 (1-401)
Qy 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTrpIleMetAlaCysMetSerAla 20
Db 3 CAARATGAATCTGCTTGACACACCCATCAAAATACATCATGGATGTCAGCT 62
Qy 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyValLeuAlaLeuAla 40
Db 63 GATCTGGAAGTAACACACAGCACCTGGGTTTGTGGAGGGTCTCGCGCCCTAGCG 122
Qy 41 AlaTyrCysLeuSerValGlyCysValIleValIleValGlyHisIleGluLeuGlyGlyLys 60
Db 123 GCCTACTGCTGTGTCAGTCGGTGTGTGTGATTGTTGGTTCATATCGAGCTGGGGGCAAG 182
Qy 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGluCys 80
Db 183 CCGGCATCTGTCACAGAAAGAGGTGTGTATCAACATACATGATGATGCAAGATGTC 242
Qy 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100

```

```

Db 243 TCACAGCTGCCCATATATGCAACAGCTCAGGTATATAGTCACCACTTCAAGGAAAA 302
Qy 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValThr 120
Db 303 GTCTTGGATTGCTGCAGCAGCACCACCAACAAGCTGTCTATTGAGCCCATAGTAAC 362
Qy 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
Db 363 ACCAACTGGCAAAAGCTTGAGGCCTTTTGGCACAAGCAT 401

RESULT 2
AAQ78041
ID AAQ78041 standard; cDNA; 401 BP.
XX
AC AAQ78041;
XX
DT 25-MAR-2003 (updated)
DT 01-AUG-1995 (first entry)
XX
DE Hepatitis C virus NS3/NS4 region.
XX
KW Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
KW classification; immunisation; prophylaxis; serotyping; ss.
XX
OS Hepatitis C virus type 3.
XX
FH Key Location/Qualifiers
FT CDS 3..401
FT /*tag= a
FT /product= NS3/NS4 polypeptide.
XX
PN W09425601-A2.
XX
PD 10-NOV-1994.
XX
PF 27-APR-1994; 94WO-EP01323.
XX
PR 27-APR-1993; 93EP-0401099.
XX
PR 05-AUG-1993; 93EP-0402019.
XX
PA (INNO-) INNOGENETICS NV SA.
XX
PI Maertens G, Stuyver L;
XX
DR WPI; 1994-358277/44.
DR P-PSDB; AAR63289.
XX
New polynucleotide sequences from hepatitis C virus - and related
vectors, polypeptide(s) and antibodies, useful for immunisation,
treatment, diagnosis and typing of HCV isolates
XX
Claim 2; Page 126; 404pp; English.
XX
Compositions comprising at least 5, and pref. 8 or more contiguous
nucleotides selected from an HCV type 3 genomic sequence, more
particularly (i) the region spanning positions 417-957 of the
Core/E1 region of HCV subtype 3a; (ii) the region spanning positions
4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
region spanning positions 8023-8235 of the NS5 region of the BR36
subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
sequence, may be used as primers to amplify nucleic acid from an
isolate belonging to a specific genotype, or as a probe for specific
detection/classification of nucleic acid. Polypeptides encoded by
the nucleotides in such compositions may be used for immunisation
against HCV, for the detection of antibodies directed against HCV
and for serotyping. This sequence corresponds to the NS3/NS4
region of HCV subtype 3a and is taken from a clone designated
BR36-20-166.
XX
(Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 401 BP; 110 A; 100 C; 101 G; 90 T; 0 other;

```


Alignment Scores:
 Pred. No.: 2,76e-77 Length: 401
 Score: 691.00 Matches: 132
 Percent Similarity: 99.25% Conservativity: 0
 Best Local Similarity: 99.25% Mismatches: 1
 Query Match: 98.86% Indels: 0
 DB: 15 Gaps: 0

US-09-638-693-36 (1-133) x AAQ78041 (1-401)

QY 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
 Db 3 CAAATGAATCTGCTTGACACACCCATCACAATAATACATCGCATGTCAGCT 62
 QY 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyValLeuAlaLeuAla 40
 Db 63 GATCTGGAAGTAACACACACCGCTGGGTTTGTCTGGAGGGTCTCGCGCCCTAGCG 122
 QY 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyLys 60
 Db 123 GCCTACTGCTGCTAGTCGGTGTGTTGATGTTGGTTCATATCGACCTGGGGGCAAG 182
 QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluCys 80
 Db 183 CCGGCAATCGTTCACAGACAAAGAGGTGTGTATCAACAATACGATGAGATGGAAGTGC 242
 QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysLys 100
 Db 243 TCACAAGTGCCCATATATCGAACAGCTCAGGTGATAGCTCACCAGTTCGAAGAAAAA 302
 QY 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValThr 120
 Db 303 GTCCCTGGATTGCTCAGGAGCCACCAACACAAAGCTGTCTATGAGCCCATAGTA 362
 QY 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
 Db 363 ACCAACTGGCAAAAGCTTGAGGCCCTTTTGGCACAAAGCAT 401

RESULT 3

AAQ78042
 ID AAQ78042 standard; cDNA; 401 BP.
 AC AAQ78042;
 DT 25-MAR-2003 (updated)
 DT 01-AUG-1995 (first entry)
 DE Hepatitis C virus NS3/NS4 region.
 XX
 KW Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
 KW classification; immunisation; prophylaxis; serotyping; ss.
 XX
 OS Hepatitis C virus type 3.
 XX
 FH Key Location/Qualifiers
 CDS 3..401
 FT /*tag= a
 FT /product= NS3/NS4 polypeptide.
 XX

PN W09425601-A2.

PD 10-NOV-1994.

XX 27-APR-1994; 94WO-EP01323.

XX 27-APR-1993; 93EP-0401099.

PR 05-AUG-1993; 93EP-0402019.

XX (INNO-) INNOGENETICS NV SA.

XX Maertens G, Stuyver L;

XX

DR WPI; 1994-358277/44.
 DR P-PSDB; AAR63290.

XX New polynucleotide sequences from hepatitis C virus - and related
 PT vectors, polypeptide(s) and antibodies, useful for immunisation,
 PT treatment, diagnosis and typing of HCV isolates
 XX
 PS
 PS
 Claim 2; Page 128; 404pp; English.

XX Compositions comprising at least 5, and pref. 8 or more contiguous
 CC nucleotides selected from an HCV type 3 genomic sequence, more
 CC particularly (i) the region spanning positions 417-957 of the
 CC Core/E1 region of HCV subtype 3a; (ii) the region spanning positions
 CC 4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
 CC positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
 CC region spanning positions 8023-8235 of the NS5 region of the BR36
 CC subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
 CC sequence, may be used as primers to amplify nucleic acid from an
 CC isolate belonging to a specific genotype, or as a probe for specific
 CC detection/classification of nucleic acid. Polypeptides encoded by
 CC the nucleotides in such compositions may be used for immunisation
 CC against HCV, for the detection of antibodies directed against HCV
 CC and for serotyping. This sequence corresponds to the NS3/NS4
 CC region of HCV subtype 3a and is taken from a clone designated
 CC BR36-20-165.
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 401 BP; 111 A; 100 C; 100 G; 90 T; 0 other;

Alignment Scores:

Pred. No.: 2,76e-77 Length: 401
 Score: 691.00 Matches: 132
 Percent Similarity: 99.25% Conservativity: 0
 Best Local Similarity: 99.25% Mismatches: 1
 Query Match: 98.86% Indels: 0
 DB: 15 Gaps: 0

US-09-638-693-36 (1-133) x AAQ78042 (1-401)

QY 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
 Db 3 CAAATGAATCTGCTTGACACACCCATCACAATAATACATCGCATGTCAGCT 62
 QY 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyValLeuAlaLeuAla 40
 Db 63 GATCTGGAAGTAACACACACCGCTGGGTTTGTCTGGAGGGTCTCGCGCCCTAGCG 122
 QY 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyLys 60
 Db 123 GCCTACTGCTGCTAGTCGGTGTGTTGATGTTGGTTCATATCGACCTGGGGGCAAG 182
 QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluCys 80
 Db 183 CCGGCAATCGTTCACAGACAAAGAGGTGTGTATCAACAATACGATGAGATGGAAGTGC 242
 QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysLys 100
 Db 243 TCACAAGTGCCCATATATCGAACAGCTCAGGTGATAGCTCACCAGTTCGAAGAAAAA 302
 QY 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValThr 120
 Db 303 GTCCCTGGATTGCTCAGGAGCCACCAACAAAGCTGTCTATGAGCCCATAGTA 362
 QY 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
 Db 363 ACCAACTGGCAAAAGCTTGAGGCCCTTTTGGCACAAAGCAT 401

RESULT 4

AAQ78125

ID AAQ78125 standard; cDNA; 629 BP.

XX

AC AAQ78125;

XX

DT	25-MAR-2003	(updated)
DT	18-AUG-1995	(first entry)
XX		HCV sequence used as hybridisation probe in identification method.
XX		Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
KW		classification; immunisation; prophylaxis; serotyping; ss.
XX		Hepatitis C virus.
XX		
XX		Location/Qualifiers
FT	CDS	3..629
FT		/*tag= a
FT	mat_peptide	3..629
FT		/*tag= b
XX		
XX	WO9425601-A2.	
XX		
XX	10-NOV-1994.	
XX		
XX	27-APR-1994;	94WO-EP01323.
XX		
XX	27-APR-1993;	93EP-0401099.
PR		
XX	05-AUG-1993;	93EP-0402019.
PR		
XX		
XX	(INNO-)	INNOGENETICS NV SA.
PA		
XX		
XX	Maertens G,	Stuyver L;
PI		
PI		
XX		
DR	WPI; 1994-358277/44.	
DR	P-PSDB; AAR63390.	
XX		
PT		New polynucleotide sequences from hepatitis C virus - and related
PT		vectors, polypeptide(s) and antibodies, useful for immunisation,
PT		treatment, diagnosis and typing of HCV isolates
XX		
PS	Disclosure; Page 274-275;	404pp; English.
XX		
XX		Compositions comprising at least 5, and pref. 8 or more contiguous
CC		nucleotides selected from an HCV type 3 genomic sequence, more
CC		particularly (i) the region spanning positions 417-957 of the
CC		Core/E1 region of HCV subtype 3a; (ii) the region spanning positions
CC		4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
CC		positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
CC		region spanning positions 8023-8235 of the NS5 region of the BR36
CC		subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
CC		sequence, may be used as primers to amplify nucleic acid from an
CC		isolate belonging to a specific genotype, or as a probe for specific
CC		detection/classification of nucleic acid. Polypeptides encoded by
CC		the nucleotides in such compositions may be used for immunisation
CC		against HCV, for the detection of antibodies directed against HCV
CC		and for serotyping.
CC		(Updated on 25-MAR-2003 to correct PN field.)
XX		
SQ	Sequence 629 BP; 159 A; 168 C; 156 G; 146 T; 0 other;	
	Alignment Scores:	
	Pred. No.:	5, 08e-77 Length: 629
	Score:	691.00 Matches: 132
	Percent Similarity:	99.25% Conservative: 0
	Best Local Similarity:	99.25% Mismatches: 1
	Query Match:	98.86% Indels: 0
	DB:	15 Gaps: 0
US-09-638-693-36	(1-133) x	AAQ78125 (1-629)
QY	1	GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
DB	231	CAAAATGAATCTGCTTGACACACCCCATCACAAATACATCATGCATGTCAGCT 290
QY	21	AspLeuGluValThrThrSerThrTrpValLeuLeuGlyValLeuAlaLeuAla 40
DB	291	GATCTGGAAGTAAACACACAGACACCTGGGTTTGTCTTGAGGGGTCTCGCGGCCCTAGCG 350

CC isolate belonging to a specific genotype, or as a probe for specific
 CC detection/classification of nucleic acid. Polypeptides encoded by
 CC the nucleotides in such compositions may be used for immunisation
 CC against HCV, for the detection of antibodies directed against HCV
 CC and for serotyping. This sequence corresponds to the NS3/NS4
 CC region of HCV subtype 3a and is taken from a clone designated
 CC HD10-1-25.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 401 BP; 106 A; 108 C; 104 G; 83 T; 0 other;

Alignment Scores:
 Pred. No.: Length: 401
 Score: 672.00 Matches: 124
 Percent Similarity: 97.74% Conservatives: 6
 Best Local Similarity: 93.23% Mismatches: 3
 Query Match: 96.14% Indels: 0
 DB: 15 Gaps: 0

US-09-638-693-36 (1-133) x AAQ78038 (1-401)

QY 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
 Db 3 CAAATGAATCTGCTTGACACACCCCGTCACAAATACATTATGCGCATGTCAGCT 62
 QY 21 AspLeuGluValThrThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLeuAla 40
 Db 63 GATCTGGAAGTAACACACACACCTGGGTGCTTGGAGGGGTCTCGCGCCCTAGCG 122
 QY 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleLeuLeuGlyLys 60
 Db 123 GCCTACTGCTTGCAGTCGGCTGCTGTAATCGTGGGTGCATATCGAGCTGGGGGCAAG 182
 QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnInTyrAspGluMetGluGlyCys 80
 Db 183 CCGGCACCTGTTCCAGCAAGAGGTGTTGTATCAACAGTAGCATGATGAGAGGTGC 242
 QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPhelyLys 100
 Db 243 TCGCAAGCGCCCATACATCCACACAGCTCAGGTAAATAGCCACCATGTTCAAGGAGAAA 302
 QY 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValThr 120
 Db 303 ATCTTGGACTGCTGAGCGAGCCACCCCAACAACAGCTGTCTATTGAGCCCGCTAATAGCT 362
 QY 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
 Db 363 TCCAACTGCAAAAGCTTGAAACCTTCTGGCACAAGCAT 401

RESULT 6

AAQ78039
 ID AAQ78039 standard; cDNA; 401 BP.
 XX
 AC AAQ78039;
 XX
 DT 25-MAR-2003 (updated)
 DT 01-AUG-1995 (first entry)
 XX
 DE Hepatitis C virus NS3/NS4 region.
 XX
 KW Hepatitis C virus; HCV; primer; probe; detection; diagnosis;
 KW classification; immunisation; prophylaxis; serotyping; ss.
 XX
 OS Hepatitis C virus type 3.
 XX
 FH Key Location/Qualifiers
 FT CDS 3..401
 FT /*tag= a
 FT /product= NS3/NS4 polypeptide.
 FT
 XX
 PN W09425601-A2.
 XX
 PD 10-NOV-1994.

XX 27-APR-1994; 94WO-EP01323.
 PF
 XX 27-APR-1993; 93EP-0401099.
 PR
 05-AUG-1993; 93EP-0402019.
 XX
 PA (INNO-) INNOGENETICS NV SA.
 XX
 PI Maertens G, Stuyver L;
 XX
 DR WPI; 1994-358277/44.
 DR P-PSDB; AAR63287.
 XX
 PT New polynucleotide sequences from hepatitis C virus - and related
 PT vectors, polypeptide(s) and antibodies, useful for immunisation,
 PT treatment, diagnosis and typing of HCV isolates
 XX
 PS Claim 2; Page 122-123; 404pp; English.
 CC
 CC Compositions comprising at least 5, and pref. 8 or more contiguous
 CC nucleotides selected from an HCV type 3 genomic sequence, more
 CC particularly (i) the region spanning positions 417-957 of the
 CC Core/E1 region of HCV subtype 3a; (ii) the region spanning positions
 CC 4664-4730 of the NS3 region of HCV type 3; (iii) the region spanning
 CC positions 4892-5292 of the NS3/4 region of HCV type 3; (iv) the
 CC region spanning positions 8023-8235 of the NS5 region of the BR36
 CC subgroup of HCV subtype 3a; or (v) an HCV subtype 3c genomic
 CC sequence, may be used as primers to amplify nucleic acid from an
 CC isolate belonging to a specific genotype, or as a probe for specific
 CC detection/classification of nucleic acid. Polypeptides encoded by
 CC the nucleotides in such compositions may be used for immunisation
 CC against HCV, for the detection of antibodies directed against HCV
 CC and for serotyping. This sequence corresponds to the NS3/NS4
 CC region of HCV subtype 3a and is taken from a clone designated
 CC HD10-1-3.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 401 BP; 106 A; 108 C; 104 G; 83 T; 0 other;

Alignment Scores:
 Pred. No.: Length: 401
 Score: 672.00 Matches: 124
 Percent Similarity: 97.74% Conservatives: 6
 Best Local Similarity: 93.23% Mismatches: 3
 Query Match: 96.14% Indels: 0
 DB: 15 Gaps: 0

US-09-638-693-36 (1-133) x AAQ78039 (1-401)

QY 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
 Db 3 CAAATGAATCTGCTTGACACACCCCGTCACAAATACATTATGCGCATGTCAGCT 62
 QY 21 AspLeuGluValThrThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLeuAla 40
 Db 63 GATCTGGAAGTAACACACACACCTGGGTGCTTGGAGGGGTCTCGCGCCCTAGCG 122
 QY 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyLys 60
 Db 123 GCCTACTGCTTGCAGTCGGCTGCTGTAATCGTGGGTGCATATCGAGCTGGGGGCAAG 182
 QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnInTyrAspGluMetGluGlyCys 80
 Db 183 CCGGCACCTGTTCCAGCAAGAGGTGTTGTATCAACAGTAGCATGATGAGAGGTGC 242
 QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPhelyLys 100
 Db 243 TCGCAAGCGCCCATACATCCACACAGCTCAGGTAAATAGCCACCATGTTCAAGGAGAAA 302
 QY 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValThr 120
 Db 303 ATCTTGGACTGCTGAGCGAGCCACCCCAACAACAGCTGTCTATTGAGCCCGCTAATAGCT 362

CC HCV genome revealed the existence of three distinct groups of HCV.
CC Analysis of the region encompassing -255 to -62 of the 5' non coding
CC region (see AAQ43058-75) showed a difference of 9-14% in the nucleotide
CC sequences between the three groups. Two of the groups identified were
CC similar to those of HCV variants termed type 1 and 2, whilst the third
CC appeared to represent a novel type of virus. Comparison of the NS3
CC region (see AAR37927-30) showed a high degree of sequence diversity with
CC type 3 being phylogenetically different to type 1 and 2. The same
CC degree differentiation was noted in the NS-5 (see AAR37923-26) and core
CC region between type 3 and type 1 sequences.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 367 BP; 96 A; 92 C; 101 G; 78 U; 0 other;

```

XX 21-NOV-1991; 91GB-0024696.
XX 24-JUN-1992; 92GB-0013362.
XX (COMM-) COMMON SERVICES AGENCY.
XX
XX Chan S, Simmonds P, Yap PL;
XX WPI; 1993-182554/22.
XX P-PSDB; AAR37936.
XX
XX DNA encoding antigenic peptide(s) of new types of hepatitis C
XX virus - for diagnosing and treating HCV infection, screening
XX blood samples and identifying different HCV types
XX
XX Disclosure; Fig 9a; 120pp; English.
XX
XX The sequences given in AAO43106-111 show bases 4911-5277 of the NS4
XX region of hepatitis C virus-3 (HCV-3) samples from 5 blood donors
XX and a consensus sequence. Analysis of this and other regions of the
XX HCV genome revealed the existence of three distinct groups of HCV.
XX Analysis of the region encompassing -255 to -62 of the 5' non coding
XX region (see AAO43058-75) showed a difference of 9-14% in the nucleotide
XX sequences between the three groups. Two of the groups identified were
XX similar to those of HCV variants termed type 1 and 2, whilst the third
XX appeared to represent a novel type of virus. Comparison of the NS3
XX region (see AAR37927-30) showed a high degree of sequence diversity with
XX type 3 being phylogenetically different to type 1 and 2. The same
XX degree of differentiation was noted in the NS-5 (see AAR37923-26) and core
XX region between type 3 and type 1 sequences.
XX (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 363 BP; 90 A; 95 C; 102 G; 76 U; 0 other;
XX
XX Alignment Scores:
XX Pred. NO.: 8.06e-67 Length: 363
XX Score: 607.00 Matches: 117
XX Percent Similarity: 98.35% Conservative: 2
XX Best Local Similarity: 96.69% Mismatches: 2
XX Query Match: 86.84% Indels: 0
XX Db: 14 Gaps: 0
XX
XX US-09-638-693-36 (1-123) x AAO43110 (1-363)

```

Qy	48	CysValValIleValGlyHisIleGluLeuGlyGlyLysProAlaIleValProAspLys	67
Db	121	UGCGUUGUCAUUGGGUCAUAUGAGCUGGGAGGCAAGCCGGCACUCUGUCCAGACAAG	180
Qy	68	GluValLeuTrpGlnGlnTyraSpGluMetGluGluCysSerGlnAlaAlaProTrpIle	87
Db	181	GAGGUGUGUAUACAACAUAUCUAUGAGAUGGAGGAGUCUCGCAAGCCGCCCAUAUAUC	240
Qy	88	GluGlnAlaGlnValIleAlaHisGlnPhetyLysGlyLysValLeuGlyLeuLeuGlnArg	107
Db	241	GAACAAGCUCAGGUAUAGGCCACCCAGUUCAAGGAGAAAGUCCUUGGGUUGCUGCAGCGA	300
Qy	108	AlaThrGlnGlnGlnAlaValIleGluProIleValThrThrAspTrpGlnLysLeuGlu	127
Db	301	GCCACCCCAACAACAGGCGUGUAUGAGGCCCAUAGUAGCUACCCACUUGGCCAAAGGCUUGAG	360
Qy	128	Ala 128	
Db	361	GUU 363	

RESULT 10

PD 27-MAY-1993.
 XX
 PF 20-NOV-1992; 92WO-GB02143.
 XX
 PR 21-NOV-1991; 91GB-0024696.
 PR 24-JUN-1992; 92GB-0013362.
 XX
 PA (COMM-) COMMON SERVICES AGENCY.
 XX
 PI Chan S, Simmonds P, Yap PL;
 XX
 DR WPI; 1993-182554/22.
 DR P-PSDB; AAR37935.
 XX
 PT DNA encoding antigenic peptide(s) of new types of hepatitis C
 PT virus - for diagnosing and treating HCV infection, screening
 PT blood samples and identifying different HCV types
 XX
 PS Disclosure; Fig 9a; 120pp; English.
 XX
 CC The sequences given in AAQ43106-111 show bases 4911-5277 of the NS4
 CC region of hepatitis C virus-3 (HCV-3) samples from 5 blood donors
 CC and a consensus sequence. Analysis of this and other regions of the
 CC HCV genome revealed the existence of three distinct groups of HCV.
 CC Analysis of the region encompassing -255 to -62 of the 5' non coding
 CC region (see AAQ43058-75) showed a difference of 9-14% in the nucleotide
 CC sequences between the three groups. Two of the groups identified were
 CC similar to those of HCV variants termed type 1 and 2, whilst the third
 CC appeared to represent a novel type of virus. Comparison of the NS3
 CC region (see AAR37927-30) showed a high degree of sequence diversity with
 CC type 3 being phylogenetically different to type 1 and 2. The same
 CC degree of differentiation was noted in the NS-5 (see AAR37923-26) and core
 CC region between type 3 and type 1 sequences.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 SQ Sequence 353 BP; 91 A; 93 C; 94 G; 75 U; 0 other;

Alignment Scores:
 Pred. No.: 7,89e-64 Length: 353
 Score: 583.00 Matches: 112
 Percent Similarity: 97.44% Conservative: 2
 Best Local Similarity: 95.73% Mismatches: 3
 Query Match: 83.40% Indels: 0
 DB: 14 Gaps: 0

US-09-638-693-36 (1-133) x AAQ43109 (1-353)

Oy 8 HisProIleThrLysTyrIleMetAlaCysMetSerAlaAspLeuGluValThrSer 27
 Db 1 CACCCCAUCCGGAUAUACCCUAGGCAUGGCAUGAUGCAGCUGAUGAAGCAACCCAGC 60
 Oy 28 ThrTrpValLeuLeuGlyGlyValLeuAlaAlaLeuAlaAlaTyrCysLeuSerValcIy 47
 Db 61 ACCUGGUGUGUGUGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
 Oy 48 CysValValIleValGlyHisIleGluLeuGlyGlyLysProAlaIleValProAspLys 67
 Db 121 UGCGUGUGAUGUGGUGAUAUAGCUGGCGGGGAGCAGCAGCAGCAGCAGCAGCAGCAG 180
 Oy 68 GluValLeuTyrGlnGlnTyrAspGluMetGluCysSerGlnAlaAlaProTyrIle 87
 Db 181 GAGGUGUGUGAUAACAUAUACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
 Oy 88 GluGlnAlaGlnValIleAlaHisGlnPheLysGlyLysValLeuGlyLeuGlnArg 107
 Db 241 GAACAACUCACGUAUAGCCACCAUUAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
 Oy 108 AlaThrGlnGlnGlnAlaValIleGluProIleValThrThrAsnTrpGln 124
 Db 301 GCCACCCCAACAACGAGCUGAUGAGCCCAUAGUAGUAGUAGUAGUAGUAGUAGUAGUAG 351

RESULT 14
 AAV59361

ID AAV59361 standard; cDNA; 9646 BP.
 XX
 AC AAV59361;
 XX
 DT 11-JAN-1999 (first entry)
 XX
 DE Hepatitis C virus H77 consensus sequence.
 XX
 KW HCV; therapy; diagnosis; vector; gene therapy; vaccine; ds.
 XX
 OS Hepatitis C virus isolate H77.
 XX
 FH Key Location/Qualifiers
 FT 5'UTR 1..342
 FT /*tag= a
 FT /note= "nucleotides 1-7 are conserved (Claim 7)"
 FT CDS 362..9377
 FT /*tag= b
 FT /product= HCV polyprotein
 FT 9378..9646
 FT /*tag= c
 FT /note= "nucleotides 9549-9646 are conserved
 FT (Claim 8)"
 FT
 PN WO9839031-Al.
 XX
 PD 11-SEP-1998.
 XX
 PF 26-FEB-1998; 98WO-US04428.
 XX
 PR 04-MAR-1997; 97US-0811566.
 XX
 PA (UNIW) UNIV WASHINGTON.
 XX
 PI Kolykhalov AA, Rice CM;
 XX
 DR WPI; 1998-520770/44.
 DR P-PSDB; AAW77397.
 XX
 CC New hepatitis C virus nucleic acid clones - comprising a 5'-terminal
 CC conserved sequence, an open reading frame encoding functional
 CC components and a 3'-terminal conserved sequence
 CC
 PS Claim 6; Page 98-104; 209pp; English.
 XX
 CC This is a nucleotide sequence of hepatitis C virus (HCV) isolate
 CC H77 cDNA. It is a composite of sequenced clones isolated from a
 CC combinatorial library and a database sequence. A genetically
 CC engineered HCV nucleic acid clone is claimed which comprises, from
 CC 5' to 3' on the positive-sense nucleic acid, a functional 5',
 CC non-translated region (NTR) comprising an extreme 5'-terminal
 CC conserved sequence (see AAV59362), an open reading frame (ORF)
 CC encoding at least a portion of an HCV polyprotein whose cleavage
 CC products form functional components of HCV virus particles and RNA
 CC replication machinery, and a 3' NTR comprising an extreme
 CC 3'-terminal conserved sequence (see AAV59364), or a derivative
 CC selected from an adapted virus, live-attenuated virus, replication
 CC competent non-infectious virus, and defective virus. The HCV
 CC nucleic acid of the claim comprises at least a functional portion
 CC of the HCV H77 consensus sequence. The invention relates to the
 CC determination of functional HCV genomic RNA sequences, to
 CC construction of infectious HCV DNA clones, and to the use of the
 CC clones, or their derivatives, in therapeutic, vaccine and
 CC diagnostic applications. The invention is also directed to HCV
 CC vectors, e.g. for gene therapy or gene vaccines. The products and
 CC methods can also be used for identifying cell lines or animals that
 CC are permissive for infection with HCV, for studying HCV infection,
 CC isolating functional components of HCV, and for screening for
 CC agents capable of modulating HCV replication in vitro and in vivo.
 XX
 SQ Sequence 9646 BP; 1889 A; 2893 C; 2724 G; 2140 T; 0 other;

Alignment Scores:

Pred. No.: 2.28e-56 Length: 9646
 Score: 539.00 Matches: 101
 Percent Similarity: 83.46% Conservative: 10
 Best Local Similarity: 75.94% Mismatches: 22
 Query Match: 77.11% Indels: 0
 DB: 19 Gaps: 0

US-09-638-693-36 (1-133) x AAV59361 (1-9646)

QY 1 GlnAsnGluLeuCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
 DB 5235 CAGAAATGAATCACCCTGACGACCAATCACCATAATACATCATGATGTCGGCC 5294
 QY 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyGlyValLeuAlaAlaLeuAla 40
 DB 5295 GACCTGGAGTCTCAGACGACCTGGTCTGTTGGCGCTCTGGTCTCTGGCC 5354
 QY 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisLeuLeuGlyLys 60
 DB 5355 GCGTATTGCTGTCAACAGGCTCGTGTGTATAGTGGCAGGATTCTTGTCCGGGAG 5414
 QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluCys 80
 DB 5415 CCGCAATATACCTGACAGGAGTCTCTACAGAGTTCGATGATGAGTGGAGAGTGC 5474
 QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100
 DB 5475 TCTCAGACATTACCGTACATCGACAGGAGTCTCTGCTGACGAGTTCAGCAGAG 5534
 QY 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValThr 120
 DB 5535 GCCTCTGGCTCTGACAGCGCTCCGCGCAGCAGAGGTTATACACCCCTGCTGCCAG 5594
 QY 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
 DB 5595 ACCAACTGGCAGAACTCGAGGTTCTTCTGGCGGAGGAC 5633

RESULT 15

ABK87285
 ID ABK87285 standard; cDNA; 9646 BP.
 XX

AC ABK87285;

XX 24-SEP-2002 (first entry)

XX cDNA encoding hepatitis C virus (HCV) H77 consensus protein.

XX Hepatitis C virus; HCV; productive replication; infection; antiviral;

XX H77; consensus protein; gene; ss.

XX Hepatitis C virus.

XX Key Location/Qualifiers

FT CDS 342..9477

FT /*tag= a

FT /partial

FT /product= "Hepatitis C virus consensus protein HCV H77"

FT /transl_except= (pos:9475..9477, aa:Glx)

FT /note= "This sequence lacks a stop codon"

XX US6392028-B1.

XX 21-MAY-2002.

XX 04-MAR-1998; 98US-0034756.

XX 04-MAR-1997; 97US-039843P.

XX 04-MAR-1997; 97US-0811566.

XX (UNIW) UNIV WASHINGTON.

XX Rice CM, Kolykhalov AA;

XX

DR WPI; 2002-478540/51.
 DR P-PSDB; AAU99289.

XX New DNA or RNA comprising a hepatitis C virus sequence that contains 3'
 PT and 5' non-translated regions flanking a polyprotein coding region,
 PT useful for detecting and developing treatment for infection by
 PT hepatitis C virus -

XX Example 4; Column 65-74; 114pp; English.

XX The present invention relates to a new DNA or RNA comprising a
 CC hepatitis C virus (HCV) sequence capable of productive replication
 CC in a host cell. The molecules of the invention comprise 5' to 3' on
 CC the positive sense strand a functional HCV 5' non-translated region
 CC comprising an extreme 5'-terminal conserved sequence, an HCV polyprotein
 CC coding region and a functional HCV 3' non-translated region comprising
 CC an extreme 3'-terminal conserved region. The DNA or RNA is used to
 CC detect and develop treatment for infection by hepatitis C virus. The
 CC present nucleic acid sequence encodes the hepatitis C virus (HCV) H77
 CC consensus protein.

XX SQ Sequence 9646 BP; 1889 A; 2893 C; 2724 G; 2140 T; 0 other;

Alignment Scores:

Pred. No.: 2.28e-56 Length: 9646
 Score: 539.00 Matches: 101
 Percent Similarity: 83.46% Conservative: 10
 Best Local Similarity: 75.94% Mismatches: 22
 Query Match: 77.11% Indels: 0
 DB: 24 Gaps: 0

US-09-638-693-36 (1-133) x ABK87285 (1-9646)

QY 1 GlnAsnGluLeuCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
 DB 5235 CAGAAATGAATCACCCTGACGACCAATCACCATAATACATCATGATGTCGGCC 5294
 QY 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyGlyValLeuAlaAlaLeuAla 40
 DB 5295 GACCTGGAGTCTCAGACGACCTGGTCTGTTGGCGCTCTGGTCTCTGGCC 5354
 QY 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisLeuLeuGlyLys 60
 DB 5355 GCGTATTGCTGTCAACAGGCTCGTGTGTATAGTGGCAGGATTCTTGTCCGGGAG 5414
 QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluCys 80
 DB 5415 CCGCAATATACCTGACAGGAGTCTCTACAGAGTTCGATGATGAGTGGAGAGTGC 5474
 QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100
 DB 5475 TCTCAGACATTACCGTACATCGACAGGAGTCTCTGCTGACGAGTTCAGCAGAG 5534
 QY 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValThr 120
 DB 5535 GCCTCTGGCTCTGACAGCGCTCCGCGCAGCAGAGGTTATACACCCCTGCTGCCAG 5594
 QY 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
 DB 5595 ACCAACTGGCAGAACTCGAGGTTCTTCTGGCGGAGGAC 5633

Search completed: August 29, 2003, 11:30:13

Job time : 245 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 29, 2003, 11:25:30 ; Search time 58 seconds

(without alignments)
1012.137 Million cell updates/sec

Title: US-09-638-693-36

Perfect score: 699

Sequence: 1 QNEICLTHPTIKYIMACMSA.....VIEPIVTTNWKLEAFWKKH 133

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q/cgn2_1/USPTO.spool/US09638693/runat_25082003_173300_21671/app_query.fasta_1.327
-DB-Issued Patents_NA -OFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOFC=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09638693 @cgn_1_1_56 @runat_25082003_173300_21671 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents_NA.*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*

2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*

3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*

4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*

5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	572	81.8	367	1	US-08-244-116B-16
2	539	77.1	9646	3	Sequence 16, Appl
3	539	77.1	9646	4	Sequence 1, Appl
4	539	77.1	12980	3	Sequence 5, Appl
5	539	77.1	12980	4	Sequence 5, Appl
6	534	76.4	943	2	Sequence 43, Appl
7	534	76.4	943	2	Sequence 43, Appl
8	534	76.4	943	2	Sequence 43, Appl
9	534	76.4	943	3	Sequence 43, Appl
10	534	76.4	1310	3	Sequence 67, Appl
11	534	76.4	1382	3	Sequence 19, Appl
12	534	76.4	1414	3	Sequence 51, Appl

13	534	76.4	1420	3	US-08-867-611-57	Sequence 57, Appl
14	534	76.4	1791	3	US-08-867-611-15	Sequence 15, Appl
15	534	76.4	1791	5	PCT-US92-06965A-20	Sequence 20, Appl
16	534	76.4	1797	3	US-08-867-611-17	Sequence 17, Appl
17	534	76.4	1797	5	PCT-US92-06965A-22	Sequence 22, Appl
18	534	76.4	2219	3	US-08-444-818-147	Sequence 147, App
19	534	76.4	2579	3	US-08-444-818-29	Sequence 29, Appl
20	534	76.4	3075	1	US-07-910-760-11	Sequence 11, Appl
21	534	76.4	3075	4	US-08-440-519-11	Sequence 11, Appl
22	534	76.4	3075	4	US-08-440-549-11	Sequence 11, Appl
23	534	76.4	6785	3	US-08-444-818-65	Sequence 65, Appl
24	534	76.4	7310	3	US-08-444-818-74	Sequence 74, Appl
25	534	76.4	8316	3	US-08-444-818-88	Sequence 88, Appl
26	534	76.4	8987	3	US-08-444-818-137	Sequence 137, App
27	534	76.4	9185	3	US-08-444-818-122	Sequence 122, App
28	534	76.4	9185	3	US-08-444-818-123	Sequence 123, App
29	534	76.4	9379	3	US-08-444-818-176	Sequence 176, App
30	534	76.4	9379	3	US-09-388-874-1	Sequence 1, Appl
31	534	76.4	9379	4	US-09-916-359-1	Sequence 1, Appl
32	534	76.4	9401	1	US-07-910-760-9	Sequence 9, Appl
33	534	76.4	9401	1	US-08-440-519-9	Sequence 9, Appl
34	534	76.4	9401	2	US-08-432-693-1	Sequence 1, Appl
35	534	76.4	9401	4	US-08-440-549-9	Sequence 9, Appl
36	534	76.4	9401	4	US-08-823-895A-25	Sequence 25, Appl
37	534	76.4	9401	5	PCT-US91-02225-9	Sequence 9, Appl
38	534	76.4	9416	3	US-08-811-566-19	Sequence 19, Appl
39	534	76.4	9416	4	US-09-034-756-19	Sequence 19, Appl
40	534	76.4	9416	4	US-08-823-895A-26	Sequence 26, Appl
41	534	76.4	9599	3	US-09-014-416-2	Sequence 2, Appl
42	534	76.4	9599	3	US-09-014-416-6	Sequence 6, Appl
43	534	76.3	712	3	US-08-444-818-13	Sequence 13, Appl
44	529	75.7	943	2	US-08-483-695-6	Sequence 6, Appl
45	529	75.7	943	2	US-07-965-285-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-08-244-116B-16
; Sequence 16, Application US/08244116B
; Patent No. 5763159
; GENERAL INFORMATION:
; APPLICANT: Simmonds, Peter
; APPLICANT: Chan, Shiu-Wan
; APPLICANT: Yap, Peng L.
; TITLE OF INVENTION: Hepatitis-C Virus Testing
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell, Seltzer, Park & Gibson, P.A.
; STREET: 1211 East Morehead Street
; CITY: Charlotte
; STATE: No. 5763159th Carolina
; COUNTRY: United States
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0. Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/244,116B
; FILING DATE: 15-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB92/02143
; FILING DATE: 20-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 1749-125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 704-377-1561
; TELEFAX: 704-334-2014

```
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 367 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Hepatitis-C virus
;
US-08-244-116B-16

Alignment Scores:
Pred. No.: 1,26e-61 Length: 367
Score: 572.00 Matches: 111
Percent Similarity: 94.87% Conservative: 0
Best Local Similarity: 94.87% Mismatches: 6
Query Match: 81.83% Indels: 0
DB: 1 Gaps: 0

US-09-638-693-36 (1-133) x US-08-244-116B-16 (1-367)

Qy 8 HisProIleThrLysTyrIleMetAlaCysMetSerAlaAspLeuValThrSer 27
Db 5 CACCCYRUCRRAAUACVUCAGGCAUGYAUUGCAGCUGAUCUGGAUAACCCAGC 64
Qy 28 ThrTrpValLeuLeuGlyGlyValValLeuAlaAlaTyrCysLeuSerValGly 47
Db 65 ACCUGGUGUGUGCUGGAGRGUCUCGCKGCCUCAGCGGCUACUGUCUGAGUCGGC 124
Qy 48 CysValValIleValGlyHisIleGluLeuGlyLysProAlaIleValProAspLys 67
Db 125 UGCGUUGUAGUUGUGGYCAUAUUGACGUGGGRGCAAGCCVGCAMUCGUUCCAGACAAR 184
Qy 68 GluValLeuTyrGlnGlnTyrAspGluMetGluCysSerGlnAlaAlaProTyrIle 87
Db 185 GARGUGUGUUAUCAACAAUACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 244
Qy 88 GluGlnAlaGlnValIleAlaHisGlnPhelLysGlyLysValLeuGlyLeuGlnArg 107
Db 245 GAACAGCUCARGURAUAGCCACCAGUUAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 304
Qy 108 AlaThrGlnGlnGlnAlaValIleGluProIleValThrTrpAsnTrpGln 124
Db 305 GCCACCAACAACARGCUGUVAUUGAGCCMAUAGUAGCUACCAACUGGCAA 355

RESULT 2
US-08-811-566-1
; Sequence 1, Application US/08811566
; Patent No. 6127116
; GENERAL INFORMATION:
; APPLICANT: Rice, Charles et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; TITLE OF INVENTION: VIRUS (HCV) AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,566
; FILING DATE: 03-MAR-1997
; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1113-1-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9646 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-811-566-1

Alignment Scores:
Pred. No.: 1,11e-55 Length: 9646
Score: 539.00 Matches: 101
Percent Similarity: 83.46% Conservative: 10
Best Local Similarity: 75.94% Mismatches: 22
Query Match: 77.11% Indels: 0
DB: 3 Gaps: 0

US-09-638-693-36 (1-133) x US-08-811-566-1 (1-9646)

Qy 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
Db 5235 CAGAATGAAGTCACCCCTGAGCAGCACCAATCACCATAATCATCATGATGATGCGGCC 5294
Qy 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyGlyValLeuAlaAlaLeuAla 40
Db 5295 GACCTGGAGGTGCTCAGCAGCACCTGGTGTCTGCGCGCTCTGCTGCTCTGCTGCTG 5354
Qy 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyLys 60
Db 5355 GCGTATTGCTGTCAACAGAGCTGCGTGTATAGTGGGAGGATGTCTTGTCCGGGAAG 5414
Qy 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGluCys 80
Db 5415 CCGGCAATTATACCTGACAGGAGGTCTCTACAGAGATTCGATGATGAGATGGAAGATGC 5474
Qy 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPhelLysGlyLys 100
Db 5475 TCTCAGCACTTACCGTACATCGAGCAGGAGATGCTCGCTGAGCAGTTCACAGCAGAG 5534
Qy 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValThr 120
Db 5535 GCGCTCGGCTCTGAGACCGGCTCCGCGCAAGCAGAGGTTATCACCCCTGCTGTCTCCAG 5594
Qy 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
Db 5595 ACCAACTGGCAGAACTCGAGGTCTTCTGGGCGGAGAGCAG 5633

RESULT 3
US-09-034-756-1
; Sequence 1, Application US/09034756
; Patent No. 6392028
; GENERAL INFORMATION:
; APPLICANT: RICE, CHARLES et al.
; TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
; TITLE OF INVENTION: VIRUS (HCV) AND USES THEREOF
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOWELL & HAFERKAMP, L.C.
; STREET: 7733 FORSYTH BLVD., SUITE 1400
; CITY: ST. LOUIS
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; US-09-034-756-1
```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: US/09/034,756
FILING DATE: 04-May-1998
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 6029-4831
TELECOMMUNICATION INFORMATION:
TELEPHONE: 314-727-5188
TELEFAX: 314-727-6092
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9646 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-034-756-1

Alignment Scores:
Pred. No.: 1,11e-55 Length: 9646
Score: 539.00 Matches: 101
Percent Similarity: 83.46% Conservative: 10
Best Local Similarity: 75.94% Mismatches: 22
Query Match: 77.11% Indels: 0
DB: 4 Gaps: 0

US-09-638-693-36 (1-133) x US-09-034-756-1 (1-9646)

QY 1 GlnAsnGluLeuCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
Db 5235 CAGATGAAGTACCCCTGACGACCAATCACCATAATATCATGATGCGGCC 5294
QY 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyValLeuAlaLeuAla 40
Db 5295 GACCTGGAGTGTCTACGAGCAGCTGGTGTCTGGGGGCTCTGCTGCGCC 5354
QY 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleLeuGlyLys 60
Db 5355 GCGTATTGCTGTCAACAGGCTGCGTGTCTAGTGGGCGAGGATTCTTGTCCGGGAAG 5414
QY 61 ProAlaIleValProAspLysGluValLeuLeuGlnGlnTyrAspGluMetGluGlyCys 80
Db 5415 CCGGCAATTATACCTGACAGGAGGTTCTCTACAGGAGTTCGATGAGATGGAAGTGC 5474
QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPhelLys 100
Db 5475 TCTCAGCACTTACCTGACAGGAGGTTCTCTACAGGAGTTCGATGAGATGGAAGTGC 5534
QY 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleLeuProIleValThr 120
Db 5535 GCGCTGGGCTCTCTGAGACCGGCTCCCGCAAGCAGAGGTTATCACCCTGCTGTCAG 5594
QY 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
Db 5595 ACCAACTGGCAGAACTCGAGGTCTCTCTGGGGCAGAC 5633

RESULT 4

US-08-811-566-5
Sequence 5, Application US/08811566
Patent No. 6127116
GENERAL INFORMATION:
APPLICANT: Rice, Charles et al.
TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
TITLE OF INVENTION: VIRUS (HCV) AND USES THEREOF

NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
STREET: Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/811,566
FILING DATE: 03-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 1113-1-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 12980 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-811-566-5

Alignment Scores:
Pred. No.: 1,65e-55 Length: 12980
Score: 539.00 Matches: 101
Percent Similarity: 83.46% Conservative: 10
Best Local Similarity: 75.94% Mismatches: 22
Query Match: 77.11% Indels: 0
DB: 3 Gaps: 0
US-09-638-693-36 (1-133) x US-08-811-566-5 (1-12980)

QY 1 GlnAsnGluLeuCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
Db 5235 CAGATGAAGTACCCCTGACGACCAATCACCATAATATCATGATGCGGCC 5294
QY 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyValLeuAlaLeuAla 40
Db 5295 GACCTGGAGTGTCTACGAGCAGCTGGTGTCTGGGGGCTCTGCTGCTGCGCC 5354
QY 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleLeuGlyLys 60
Db 5355 GCGTATTGCTGTCAACAGGCTGCGTGTCTAGTGGGCGAGGATTCTTGTCCGGGAAG 5414
QY 61 ProAlaIleValProAspLysGluValLeuLeuGlnGlnTyrAspGluMetGluGlyCys 80
Db 5415 CCGGCAATTATACCTGACAGGAGGTTCTCTACAGGAGTTCGATGAGATGGAAGTGC 5474
QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPhelLys 100
Db 5475 TCTCAGCACTTACCTGACAGGAGGTTCTCTACAGGAGTTCGATGAGATGGAAGTGC 5534
QY 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleLeuProIleValThr 120
Db 5535 GCGCTGGGCTCTCTGAGACCGGCTCCCGCAAGCAGAGGTTATCACCCTGCTGTCAG 5594
QY 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
Db 5595 ACCAACTGGCAGAACTCGAGGTCTCTCTGGGGCAGAC 5633

Db 534 CAGATGAATACACCTGACGCCACCCAGTCACCAATATACATCATGATGTCGGCC 593

QY 21 AspLeuGluValThrThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLeuAla 40
|||||
Db 594 GACCTGGAGTCTGTCAGGACACCTGGTCTGTTGGCGGCTCTGCTGCTTTGGCC 653
|||||
QY 41 AlaTyrCysLeuSerValGlyCysValValLeuValHisLeuGlyGlyLys 60
|||||
Db 654 GCGTATGCTGTCACAGGCTCGTGGTCATAGTGGGCGGCTGCTTTGCCGGAAG 713
|||||

QY 61 ProAlaLeuValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGlyCys 80
|||||
Db 714 CCGGCAATACATCTGACAGGGAAGTCTCTACCGAGAGTTCGATGAGATGGAAGTGC 773
|||||

QY 81 SerGlnAlaAlaProTyrLeuGlnAlaGlnValLeuAlaHisGlnPheLysGlyLys 100
|||||
Db 774 TCTCAGCAGCTTACCGTACATCGAGCAAGGATGCTCGCCGAGCAGTTCACAGCAGAAG 833
|||||

QY 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValLeuGluProLeuValThr 120
|||||
Db 834 GCCTCGGCTCTGTCAGACCGCTCGCTCAGCAGAGGTTATCGCCCTGCTGTCCAG 893
|||||

QY 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
|||||

Db 894 ACCAACTGGCAAACTCGAGACCTTCTGGCGGAAGCAT 932
|||||

RESULT 7

US-07-965-285-43

; Sequence 43, Application US/07965285

; Patent No. 5879904

; GENERAL INFORMATION:

; APPLICANT: Brechot, Christian

; APPLICANT: Kremsdorf, Dina

; APPLICANT: Porchon, Colette

; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a

; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic

; TITLE OF INVENTION: Applications

; NUMBER OF SEQUENCES: 46

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

; ADDRESSEE: Dunner

; STREET: 1300 I Street, N.W.

; CITY: Washington

; STATE: DC

; COUNTRY: USA

; ZIP: 20005-3315

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07965,285

; FILING DATE: 18-MAR-1993

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; FILING DATE: 06-JUN-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Meyers, Kenneth J.

; REGISTRATION NUMBER: 25,146

; REFERENCE/DOCKET NUMBER: 05286-0001-00000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-408-4000

; TELEFAX: 202-408-4400

; INFORMATION FOR SEQ ID NO: 43:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 943 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: Other

; DESCRIPTION: cdna to genomic RNA

US-07-965-285-43

Alignment Scores: 2.03e-56 Length: 943

Pred. No.: 534.00 Matches: 99

Score: 83.46% Conservative: 12

Percent Similarity: 74.44% Mismatches: 22

Best Local Similarity: 76.39% Indels: 0

Query Match: 2 Gaps: 0

DB: 0

US-09-638-693-36 (1-133) x US-07-965-285-43 (1-943)

QY 1 GlnAsnGluLeuCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
|||||

Db 534 CAGATGAATACACCTGACGCCACCCAGTCACCAATATACATCATGATGTCGGCC 593
|||||

QY 21 AspLeuGluValThrThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLeuAla 40
|||||

Db 594 GACCTGGAGTCTGTCAGGACACCTGGTCTGTTGGCGGCTCTGCTGCTTTGGCC 653
|||||

QY 41 AlaTyrCysLeuSerValGlyCysValValLeuValHisLeuGlyGlyLys 60
|||||

Db 654 GCGTATGCTGTCACAGGCTCGTGGTCATAGTGGGCGGCTGCTTTGCCGGAAG 713
|||||

QY 61 ProAlaLeuValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGlyCys 80
|||||

Db 714 CCGGCAATACATCTGACAGGGAAGTCTCTACCGAGAGTTCGATGAGATGGAAGTGC 773
|||||

QY 81 SerGlnAlaAlaProTyrLeuGlnAlaGlnValLeuAlaHisGlnPheLysGlyLys 100
|||||

Db 774 TCTCAGCAGCTTACCGTACATCGAGCAAGGATGCTCGCCGAGCAGTTCACAGCAGAAG 833
|||||

QY 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValLeuGluProLeuValThr 120
|||||

Db 834 GCCTCGGCTCTGTCAGACCGCTCGCTCAGCAGAGGTTATCGCCCTGCTGTCCAG 893
|||||

QY 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
|||||

Db 894 ACCAACTGGCAAACTCGAGACCTTCTGGCGGAAGCAT 932
|||||

RESULT 8

US-08-487-231-43

; Sequence 43, Application US/08487231

; Patent No. 5919454

; GENERAL INFORMATION:

; APPLICANT: Brechot, Christian

; APPLICANT: Kremsdorf, Dina

; APPLICANT: Porchon, Colette

; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a

; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic

; TITLE OF INVENTION: Applications

; NUMBER OF SEQUENCES: 46

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

; ADDRESSEE: Dunner

; STREET: 1300 I Street, N.W.

; CITY: Washington

; STATE: DC

; COUNTRY: USA

; ZIP: 20005-3315

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08487,231

; FILING DATE: 07-JUNE-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; FILING DATE: 18-MAR-1993

; CLASSIFICATION: 435

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: FR 91 06 882
;; FILING DATE: 06-JUN-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Meyers, Kenneth J.
;; REGISTRATION NUMBER: 25,146
;; REFERENCE/DOCKET NUMBER: 05286-0001-02000
;; TELEPHONE: 202-408-4000
;; TELEFAX: 202-408-4400
;; INFORMATION FOR SEQ ID NO: 43:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 943 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: Other
;; DESCRIPTION: CDNA to genomic RNA
US-08-487-231-43

Alignment Scores:
Pred. No.: 2.03e-56 Length: 943
Score: 534.00 Matches: 99
Percent Similarity: 83.46% Conservative: 12
Best Local Similarity: 74.44% Mismatches: 22
Query Match: 76.39% Indels: 0
DB: 2 Gaps: 0

US-09-638-693-36 (1-133) x US-08-487-231-43 (1-943)

Qy 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
|||||
Db 534 CAGAATGAATCACCCTGACGACCCAGTACCAAAATACATCATGACATGCTCGGCC 593
Qy 21 AspleuGluValThrSerThrTrpValLeuLeuGlyValLeuAlaLeuAla 40
|||||
Db 594 GACCTGGAGGTCTGACAGACCCCTGGGTCTCTGGCGGCTGCTGCTGCTTTGGCC 653
Qy 41 AlaTyrCysLeuSerValGlyCysValIleValGlnHisIleGluLeuGlyLys 60
|||||
Db 654 GCGTATTGCTGTCAACAGGCTGGTGTATAGTGGCGAGGCTGCTGCTGCTGCTGCTG 713
Qy 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGlyLys 80
|||||
Db 714 CCGCAATCATACCTGACAGGAAGTCTCTACCGAGAGTTCGATGATGGAAGAGTGC 773
Qy 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100
|||||
Db 774 TCTCAGCATTACCGTATACGACAGGAGGTGATGCTGCCGCGAGCTTCAACGAGAAG 833
Qy 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValThr 120
|||||
Db 834 GCGCTCGGCTCTCTGACAGCCGCTCCCGTCAGGAGGTATCGCCCTGCTGCTCCAG 893
Qy 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
|||||
Db 894 ACCAACTGGCAAAACTCGAGACCTTCTGGCGGAAGCAT 932

RESULT 9

US-09-201-912-43
; Sequence 43, Application US/09201912
; Patent No. 6210962
; GENERAL INFORMATION:
; APPLICANT: Brechot, Christian
; APPLICANT: Kremsdorf, Dina
; APPLICANT: Porchon, Colette
; TITLE OF INVENTION: Nucleotide and Peptide Sequences of a
; TITLE OF INVENTION: Hepatitis C Virus Isolate, Diagnostic and Therapeutic
; TITLE OF INVENTION: Applications
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner

;; STREET: 1300 I Street, N.W.
;; CITY: Washington
;; STATE: DC
;; COUNTRY: USA
;; ZIP: 20005-3315
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent in Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/201,912
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 07/965,285
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Meyers, Kenneth J.
;; REGISTRATION NUMBER: 25,146
;; REFERENCE/DOCKET NUMBER: 05286-0001-00000
;; TELEPHONE: 202-408-4000
;; TELEFAX: 202-408-4400
;; INFORMATION FOR SEQ ID NO: 43:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 943 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: Other
;; DESCRIPTION: CDNA to genomic RNA
US-09-201-912-43

Alignment Scores:
Pred. No.: 2.03e-56 Length: 943
Score: 534.00 Matches: 99
Percent Similarity: 83.46% Conservative: 12
Best Local Similarity: 74.44% Mismatches: 22
Query Match: 76.39% Indels: 0
DB: 2 Gaps: 0

US-09-638-693-36 (1-133) x US-09-201-912-43 (1-943)

Qy 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
|||||
Db 534 CAGAATGAATCACCCTGACGACCCAGTACCAAAATACATCATGACATGCTCGGCC 593
Qy 21 AspleuGluValThrSerThrTrpValLeuLeuGlyValLeuAlaLeuAla 40
|||||
Db 594 GACCTGGAGGTCTGACAGACCCCTGGGTCTCTGGCGGCTGCTGCTGCTTTGGCC 653
Qy 41 AlaTyrCysLeuSerValGlyCysValIleValGlnHisIleGluLeuGlyLys 60
|||||
Db 654 GCGTATTGCTGTCAACAGGCTGGTGTATAGTGGCGAGGCTGCTGCTGCTGCTGCTG 713
Qy 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGlyLys 80
|||||
Db 714 CCGCAATCATACCTGACAGGAAGTCTCTACCGAGAGTTCGATGATGGAAGAGTGC 773
Qy 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100
|||||
Db 774 TCTCAGCATTACCGTATACGACAGGAGGTGATGCTGCCGCGAGCTTCAACGAGAAG 833
Qy 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValThr 120
|||||
Db 834 GCGCTCGGCTCTCTGACAGCCGCTCCCGTCAGGAGGTATCGCCCTGCTGCTCCAG 893
Qy 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
|||||
Db 894 ACCAACTGGCAAAACTCGAGACCTTCTGGCGGAAGCAT 932

RESULT 10


```

US-08-444-818-67
; Sequence 67, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1310 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; CLONE: composite of clones 36, 81, and 32
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1146
; US-08-444-818-67

Alignment Scores:
Pred. No.: 3,15e-56 Length: 1310
Score: 534.00 Matches: 99
Percent Similarity: 83.46% Conservative: 12
Best Local Similarity: 74.44% Mismatches: 22
Query Match: 76.39% Indels: 0
DB: 3 Gaps: 0

US-09-638-693-36 (1-133) x US-08-444-818-67 (1-1310)
QY 1 GlnAsnGluLeuCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
Db 232 CAGAATGAATCACCCTGACGACCCAGTCACCAATACATCATGATGTCGCGCC 291
QY 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyValLeuAlaLeuAla 40
Db 292 GACCTGGAGTGGTCTACAGACACCTGGTGGTGGCGGCTCTGCTGCTGGCC 351
QY 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyGlyLys 60
Db 352 GCGTATGCTGTCACACAGCGTCGGTGGTCATAGTGGCGAGGTCTCTTGTCCGGGAAG 411
QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGlyCys 80
Db 412 CCGGCAATCATACCTCACAGGGAAGTCCCTCTACCGAGAGTTCGATGAGATGGNAGAGTGC 471

```

```

QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100
Db 472 TCTCAGCACTTACCGTACATCGAGCAAGGATGATGCTGCCGAGCAGTTCACAGCAGAAG 531
QY 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValThr 120
Db 532 GGCCTCGGCTCTCTGAGACCGCGTCCCGTCAGCGAGAGGTATCGCCCTGCTGCTGCCAG 591
QY 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
Db 592 ACCAACTGGCAAAACTCGAGACCTTCTGGGGGAAGCAT 630

RESULT 11
US-08-444-818-19
; Sequence 19, Application US/08444818
; Patent No. 6150087
; GENERAL INFORMATION:
; APPLICANT: Chien, David Y.
; APPLICANT: Rutter, William J.
; TITLE OF INVENTION: NANBV Diagnostics and Vaccines
; NUMBER OF SEQUENCES: 777
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,818
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,590
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0110.002
; TELEPHONE: (508)359-3876
; TELEFAX: (508)359-3885
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1382 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1380
; US-08-444-818-19

Alignment Scores:
Pred. No.: 3,39e-56 Length: 1382
Score: 534.00 Matches: 99
Percent Similarity: 83.46% Conservative: 12
Best Local Similarity: 74.44% Mismatches: 22
Query Match: 76.39% Indels: 0
DB: 3 Gaps: 0

US-09-638-693-36 (1-133) x US-08-444-818-19 (1-1382)
QY 1 GlnAsnGluLeuCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
Db 484 CAGAATGAATCACCCTGACGACCCAGTCACCAATACATCATGATGTCGCGCC 543

```

```
Qy 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLeuAla 40
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 544 GACCTGGAGGTCGTCACGAGCACCTGGGTGCTGTTGGCGCGTCTGGCTGTGGCC 603
Qy 41 AlaTyrCysLeuSerValGlyCysValValleValGlyHisIleGluLeuGlyGlyLys 60
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 604 CGTATTTCCTGTCAACAGGTCGTCATAGTGGCGAGGTCGTCCTTCCGGGAAG 663
Qy 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnValTrpAspGluMetGluLys 80
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 664 CCGCAATCATACCTGACAGGAAGTCTCTACCGAGAGTTCGATGAGATGGAAGATGC 723
Qy 81 SerGlnAlaLaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 724 TCTCAGCACTTACCGTACATCGACGAGGATGATGCTCGCGAGCAGTTCACACGAG 783
Qy 101 ValLeuGlyLeuLeuGlnAlaArgAlaThrGlnGlnAlaValIleGluProIleValThr 120
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 784 GCCCTCGGCTCTCTGACAGCGCTCCGTCGACGAGGTTATCGCCCTGCTGTCCAG 843
Qy 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 844 ACCAACTGGCAAAATCGAGACCTTCTGGCGGAGCAT 882

RESULT 12
US-08-867-611-51
; Sequence 51, Application US/08867611
; Patent No. 6172189
; GENERAL INFORMATION:
; APPLICANT: DEVARE, SUSHIL G
; APPLICANT: DESAI, SURESH M
; APPLICANT: CASEY, JAMES M
; APPLICANT: DAILEY, STEPHEN H
; APPLICANT: DAWSON, GEORGE J
; APPLICANT: GUTIERREZ, ROBIN A
; APPLICANT: LESNIEWSKI, RICHARD R
; APPLICANT: STEWART, JAMES L
; APPLICANT: RUPPRECHT, KEVIN R
; TITLE OF INVENTION: HEPATITIS C ASSAY UTILIZING RECOMBINANT
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: ABBOTT LABORATORIES
; STREET: ONE ABBOTT PARK ROAD, CHAD377/AP6D2
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,611
; FILING DATE: 02-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,757
; FILING DATE:
; APPLICATION NUMBER: US/08/179,896
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/572,822
; FILING DATE: 24-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/614,069
; FILING DATE: 07-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/748,561
; FILING DATE: 21-AUG-1991
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/748,565
; FILING DATE: 21-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/748,566
; FILING DATE: 21-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKI, PRISCILLA E
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 4834.US.P6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1414 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-867-611-51

Alignment Scores:
Pred. No.: 3,49e-56 Length: 1414
Score: 534.00 Matches: 99
Percent Similarity: 83.46% Conservative: 12
Best Local Similarity: 74.44% Mismatches: 22
Query Match: 76.39% Indels: 0
DB: 3 Gaps: 0

US-09-638-693-36 (1-133) x US-08-867-611-51 (1-1414)
Qy 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 487 CAGACGAATACACCTGACCCCGGTTACCAAAATACATCATGACCTGCATGCTGCT 546
Qy 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLeuAla 40
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 547 GATCTAGAGTTGTACCTCTACCTGCGGTTCTGTTGTTGTTGTTGTTGTTGTTGTT 605
Qy 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyGlyLys 60
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 607 CTTACTGCTGTCGACCGGTTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 666
Qy 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluLys 80
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 667 CCGGCCATTATCCCGGACCGTGAAGTCTCTACCGTCAAGTTCGACGAAATGGAAGAATGC 726
Qy 81 SerGlnAlaLaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 727 TCTCAGCACCTGCGCTACATCGAACAGGGTATGATGCTGCTGAACAGATTCAACAGAAA 786
Qy 101 ValLeuGlyLeuLeuGlnAlaArgAlaThrGlnGlnAlaValIleGluProIleValThr 120
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 787 GCTCTGGGTCGTCGACAGCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 846
Qy 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
|||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 847 ACCAACTGGCAGAAATCGAGACCTTCTGGCGTAAACAC 885

RESULT 13
US-08-867-611-57
; Sequence 57, Application US/08867611
; Patent No. 6172189
; GENERAL INFORMATION:
; APPLICANT: DEVARE, SUSHIL G
; APPLICANT: DESAI, SURESH M
; APPLICANT: CASEY, JAMES M
; APPLICANT: DAILEY, STEPHEN H
; APPLICANT: DAWSON, GEORGE J
; APPLICANT: GUTIERREZ, ROBIN A
; APPLICANT: LESNIEWSKI, RICHARD R
; APPLICANT: STEWART, JAMES L
; APPLICANT: RUPPRECHT, KEVIN R
```


ATTORNEY/AGENT INFORMATION:
NAME: POREMSKI, PRISCILLA E
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 4834.US.P6
TELEPHONE: 708-937-6365
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1791 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1791
US-08-867-611-15

Alignment Scores:
Pred. No.: 4,79e-56 Length: 1791
Score: 534.00 Matches: 99
Percent Similarity: 83.46% Conservative: 12
Best Local Similarity: 74.44% Mismatches: 22
Query Match: 76.39% Indels: 0
DB: 3 Gaps: 0

US-09-638-693-36 (1-133) x US-08-867-611-15 (1-1791)

QY 1 GlnAsnGluLeuCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
Db 862 CAGAACGAAATCACCCTGACCCGCGTTACCAAAATACATCATGACCTGCTGCTGCT 921
QY 21 AspLeuGluValThrSerThrTrpValLeuGluGlyValLeuAlaLeuAla 40
Db 922 GATCTAGAAAGTTGTTACCTCTACCTGCGGTTCTGTTGGTGTCTGCTGCTGCT 981
QY 41 AlatyrcysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyLys 60
Db 982 GCTTACTGCTGTCGACCGGTTGCTGTTATCGTGTGCTGCTGCTGCTGCTGCTGCT 1041
QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGluCys 80
Db 1042 CCGGCCATTATCCCGGACCGGTAAGTTCTGTACCGTGAGTTTCGACGAAATGGAAGATGC 1101
QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100
Db 1102 TCTCAGCACCTGCGGTACATCGACAGGGTATGATGCTGGCTGAACAGTTCAACAGAAA 1161
QY 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValThr 120
Db 1162 GCTCTGGGCTGCTGACAGCGCTTCTGTCAGGCTGAAGTTATCGCTCCGGCTGTTTCAG 1221
QY 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
Db 1222 ACCAACTGGCAGAACTCGAGACCTTCTGGGCTAAACAC 1260

RESULT 15
PCT-US92-06965A-20
Sequence 20, Application PC/TUS9206965A
GENERAL INFORMATION:
APPLICANT: DEVARE, S.
APPLICANT: DESAI, S.
APPLICANT: DAILEY, S.
TITLE OF INVENTION: HCV SYNTHETIC PEPTIDE FROM NS1 REGION
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES
STREET: ONE ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: ILLINOIS
COUNTRY: U.S.
ZIP: 60065-3500

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/06965A
FILING DATE: 19920821
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: POREMSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 4834PC.02
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 1791 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1791
PCT-US92-06965A-20

Alignment Scores:
Pred. No.: 4,79e-56 Length: 1791
Score: 534.00 Matches: 99
Percent Similarity: 83.46% Conservative: 12
Best Local Similarity: 74.44% Mismatches: 22
Query Match: 76.39% Indels: 0
DB: 5 Gaps: 0

US-09-638-693-36 (1-133) x PCT-US92-06965A-20 (1-1791)

QY 1 GlnAsnGluLeuCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
Db 862 CAGAACGAAATCACCCTGACCCGCGTTACCAAAATACATCATGACCTGCTGCTGCT 921
QY 21 AspLeuGluValThrSerThrTrpValLeuGluGlyValLeuAlaLeuAla 40
Db 922 GATCTAGAAAGTTGTTACCTCTACCTGCGGTTCTGTTGGTGTCTGCTGCTGCTGCT 981
QY 41 AlatyrcysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyLys 60
Db 982 GCTTACTGCTGTCGACCGGTTGCTGTTATCGTGTGCTGCTGCTGCTGCTGCTGCT 1041
QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGluCys 80
Db 1042 CCGGCCATTATCCCGGACCGGTAAGTTCTGTACCGTGAGTTTCGACGAAATGGAAGATGC 1101
QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100
Db 1102 TCTCAGCACCTGCGGTACATCGACAGGGTATGATGCTGGCTGAACAGTTCAACAGAAA 1161
QY 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValThr 120
Db 1162 GCTCTGGGCTGCTGACAGCGCTTCTGTCAGGCTGAAGTTATCGCTCCGGCTGTTTCAG 1221
QY 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
Db 1222 ACCAACTGGCAGAACTCGAGACCTTCTGGGCTAAACAC 1260

Search completed: August 29, 2003, 12:53:23
Job time : 66 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 29, 2003, 11:26:00 ; Search time 216 seconds

(without alignments)
1412.664 Million cell updates/sec

Title: US-09-638-693-36

Perfect score: 699

Sequence: 1 ONEILTHPTIKYIMACMSA.....VIEPIVTNNQKLEAFWIKH 133

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 1533700 seqs, 1147125425 residues

Total number of hits satisfying chosen parameters: 3067400

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL-frame+ p2n.model -DEV=xlh
-Q/cgn2_1/USPTO.spool/US09638693/runat_25082003_173300_21726/app_query.fasta_1.327
-DB-PublishedApplications_NA -QMT-fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS-human40.cdi -LIST=45 -LOCAL-ALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=DOCL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USPR=US09638693 @CGN_1.1.221 @runat_25082003_173300_21726
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA:

```
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match Length	ID	Description
1	699	100.0	401 11	US-09-899-046-35 Sequence 35, Appl

2	699	100.0	401	11	US-09-878-281-35	Sequence 35, Appl
3	691	98.9	401	11	US-09-899-046-37	Sequence 37, Appl
4	691	98.9	401	11	US-09-898-046-39	Sequence 39, Appl
5	691	98.9	401	11	US-09-878-281-37	Sequence 37, Appl
6	691	98.9	401	11	US-09-878-281-39	Sequence 39, Appl
7	691	98.9	629	11	US-09-899-046-222	Sequence 222, App
8	691	98.9	629	11	US-09-878-281-222	Sequence 222, App
9	672	96.1	401	11	US-09-899-046-31	Sequence 31, Appl
10	672	96.1	401	11	US-09-899-046-33	Sequence 33, Appl
11	672	96.1	401	11	US-09-878-281-31	Sequence 31, Appl
12	672	96.1	401	11	US-09-878-281-33	Sequence 33, Appl
13	539	77.1	9646	9	US-09-742-659-3	Sequence 3, Appl
14	539	77.1	9646	10	US-09-238-076-1	Sequence 1, Appl
15	539	77.1	9646	11	US-09-995-937-1	Sequence 1, Appl
16	539	77.1	9646	11	US-09-917-563-1	Sequence 5, Appl
17	539	77.1	12980	10	US-09-238-076-5	Sequence 5, Appl
18	539	77.1	12980	11	US-09-995-937-5	Sequence 5, Appl
19	539	77.1	12980	11	US-09-917-563-5	Sequence 5, Appl
20	534	76.4	9379	9	US-09-916-359-1	Sequence 1, Appl
21	534	76.4	9416	10	US-09-238-076-19	Sequence 19, Appl
22	534	76.4	9416	10	US-09-929-955-13	Sequence 13, Appl
23	534	76.4	9416	11	US-09-995-937-13	Sequence 19, Appl
24	534	76.4	9416	11	US-09-917-563-19	Sequence 19, Appl
25	534	76.4	9416	13	US-10-104-966-13	Sequence 13, Appl
26	534	76.4	10803	10	US-09-747-419-17	Sequence 17, Appl
27	534	76.4	10803	14	US-10-259-275-17	Sequence 17, Appl
28	492	70.4	13910	11	US-09-919-901-1	Sequence 1, Appl
29	492	70.4	13910	11	US-09-919-901-8	Sequence 8, Appl
30	492	70.4	13910	11	US-09-919-901-15	Sequence 15, Appl
31	490	70.1	6189	14	US-10-259-275-41	Sequence 41, Appl
32	490	70.1	7992	13	US-10-005-469-1	Sequence 1, Appl
33	490	70.1	7992	13	US-10-005-469-2	Sequence 2, Appl
34	490	70.1	7992	13	US-10-005-469-6	Sequence 6, Appl
35	490	70.1	7995	13	US-10-005-469-3	Sequence 3, Appl
36	490	70.1	8638	12	US-10-309-561-24	Sequence 24, Appl
37	490	70.1	8638	13	US-10-029-907-24	Sequence 24, Appl
38	490	70.1	8639	12	US-10-309-561-1	Sequence 1, Appl
39	490	70.1	8639	13	US-10-029-907-1	Sequence 1, Appl
40	490	70.1	8648	12	US-10-309-561-5	Sequence 5, Appl
41	490	70.1	8648	13	US-10-029-907-5	Sequence 5, Appl
42	490	70.1	9275	14	US-10-259-275-39	Sequence 39, Appl
43	490	70.1	10690	14	US-10-125-940-1	Sequence 1, Appl
44	487	69.7	7992	13	US-10-005-469-4	Sequence 4, Appl
45	487	69.7	8638	12	US-10-309-561-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1

US-09-899-046-35
; Sequence 35, Application US/09899046
; Publication No. US20030008274A1
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: New sequences of hepatitis C virus
; TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
; NUMBER OF SEQUENCES: 270
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/899,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/362,455
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 401 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

; ; TOPOLOGY: linear ; STRANDEDNESS: single
; ; MOLECULE TYPE: cDNA ; TOPOLOGY: linear ;
; ; HYPOTHETICAL: NO ; MOLECULE TYPE: cDNA ;
; ; ANTI-SENSE: NO ; HYPOTHETICAL: NO ;
; ; IMMEDIATE SOURCE: ; ANTI-SENSE: NO ;
; ; CLONE: -BR36-20-164 ; IMMEDIATE SOURCE: ;
; ; FEATURE: ; CLONE: BR36-20-164 ;
; ; NAME/KEY: CDS ; FEATURE: ;
; ; LOCATION: 3..401 ; NAME/KEY: CDS ;
; ; US-09-899-046-35 ; LOCATION: 3..401 ;
; ; US-09-878-281-35 ;

```



```

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 401 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; CLONE: BR36-20-166
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..401
US-09-878-281-37

```

Alignment Scores:		
Pred. No.:	1,76e-90	Length:
Score:	691.00	Matches:
Percent Similarity:	99.25%	Conservative:
Best Local Similarity:	99.25%	Mismatches:
Query Match:	98.86%	Indels:
DB:	11	Gaps:
		0

US-09-638-693-36 (1-133) x US-09-878-281-37 (1-401)

QY		1	GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla	20
Db		3	CAAAATGAATCTGCTTGACACACCCCATCACAAATACATCATGCATGTCACT	62
QY		21	AspLeuGluValThrSerThrTrpValLeuLeuClyGlyValLeuAlaLeuAla	40
Db		63	GATCTGGAGGTAAACCACGACACTGGGTTTGCTTGGAGGGGTCTCGCGGCCCTTAGCG	122
QY		41	AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyGlyLys	60
Db		123	GCCTACTGCTTCAGTCGGTTGTGTATTGGGTTCATATCATCGCTGGGGGGCAGA	182
QY		61	ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGluCys	80
Db		183	CCGGCAATCGTTCCACAGACAAAGAGGTGTTGTATCAACAATACGATGAGATGGAAGATGC	242
QY		81	SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys	100
Db		243	TCAACAGCTGCCCATATATCGAACAGCTCAGGTGATGCTCACCAGTTCACAGGAAAAA	302
QY		101	ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValThr	120
Db		303	GTTCCTTGATTGCTGCAGCGAGCCACCACACACAGCTGTCAATTGAGGCCATAGTACT	362
QY		121	ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis	133
Db		363	ACCAAATGGCAAAAGCTTGAGGCCCTTTTGGCACAAGCAT	401

RESULT 6

US-09-878-281-39
; Sequence 39, Application US/09878281
; Publication No. US20030032005A1

```

39:
; INFORMATION FOR SEQ ID NO:
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 401 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;
;     MOLECULE TYPE: cDNA
;     HYPOTHETICAL: NO
;     ANTI-SENSE: NO
;
;     IMMEDIATE SOURCE:
;     CLONE: BR36-20-165
;
;     FEATURE:
;
;     NAME/KEY: CDS
;     LOCATION: 3..401
;
; OS=98-878-281-39

```

Alignment Scores:		
Pred. No.:	1.76e-90	401
Score:	691.00	132
Percent Similarity:	99.25%	Conservative: 0
Best Local Similarity:	99.25%	Mismatches: 1
Query Match:	98.86%	Indels: 0
DB:	11	Gaps: 0

US-09-638-693-36 (1-133) x US-09-878-281-39 (1-401)

Qy	1	GlnAsnGluIleCysLeuThrHisProIleThrIleLysTyrIleMetAlaCysMetSerAla	20
Db	3	CAAAATGAAATCTGCTTGACACACCCCATCAAAAATACATCATGGCATGTCAGCT	62
Qy	21	AspLeuGluValThrSerThrTrpValLeuLeuGlyGlyValLeuAlaLeuAla	40
Db	63	GATCTGGAAGTAACCAACGAGCACCCTGGGTCTTGAGGGGTCCCTGGCGGCCCTAGCG	122
Qy	41	AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyGlyLys	60
Db	123	GCTACTGCTTGCAGTCGGTGTGTGTGATTGTGGTCATATCGAGCTGGGGGGCAAG	182
Qy	61	ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGluCys	80
Db	183	CGGGCAATCGTTCCAGACAAAGAGGTGTGTATCAACAATACGATGAGATGGAAGAGTGC	242
Qy	81	SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys	100
Db	243	TCACAAGCTGCCCATATATCCAACAAGCTCAGGTAATAGCTCACCAAGTTCAAGGAAAAA	302
Qy	101	ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValThr	120
Db	303	GTCCCTTGGATTGCTGCACGAGCCACCACAACAACAGCTGTCTATTGAGCCCATAGTAAC	362
Qy	121	ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis	133
Db	363	ACCAACTGCAAAAGCTTTGAGCGCTTTTGGGCACAAGCAT	401

RESULT, T 7

US-09-899-046-222 ; Sequence 222, Application US/09899046
; Publication No. US20030008274A1

```

1 / APPLICANT:
2 /
3 / TITLE OF INVENTION: New sequences of hepatitis C virus
4 /
5 / TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.
6 /
7 / NUMBER OF SEQUENCES: 270
8 /
9 / COMPUTER READABLE FORM:
10 /
11 / MEDIUM TYPE: Floppy disk
12 /
13 / COMPUTER: IBM PC compatible
14 /
15 / OPERATING SYSTEM: PC-DOS/MS-DOS
16 /
17 / SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
18 /
19 / CURRENT APPLICATION DATA:
20 /
21 / APPLICATION NUMBER: US/09/878,281
22 /
23 / FILING DATE:
24 /
25 / PRIOR APPLICATION DATA:
26 /
27 / APPLICATION NUMBER: 08/362,455
28 /
29 / FILING DATE:

```


SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA.

;; APPLICATION NUMBER: US/09/899,046
;; FILING DATE:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/362,455
;; FILING DATE:
;; INFORMATION FOR SEQ ID NO: 31:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 401 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; IMMEDIATE SOURCE:
;; CLONE: HD10-1-25
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 3..401
US-09-899-046-31

Alignment Scores:
Pred. No.: 9.85e-88 Length: 401
Score: 672.00 Matches: 124
Percent Similarity: 97.74% Conservative: 6
Best Local Similarity: 93.23% Mismatches: 3
Query Match: 96.14% Indels: 0
DB: 11 Gaps: 0

US-09-638-693-36 (1-133) x US-09-899-046-31 (1-401)

```
Qy 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
   |||
Db 3 CAAATGAAATCTGCTTGACACACCCGTCACAAAATACATTATGGCATGCTAGCT 62
   |||

Qy 21 AspleuGluValThrSerThrTrpValLeuLeuGlyValLeuAlaLeuAla 40
   |||
Db 63 GATCTGGAAGTAACACACACACCTGGGTGTTGCTTGGAGGGGTCCTCGCGCCCTAGCG 122
   |||

Qy 41 AlatyrcysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyGlyLys 60
   |||
Db 123 GCCTACTGCTGTGTCAGTCGGCTGGTGTAAATCGTGGTTCATATCGAGCTGGGGGCAAG 182
   |||

Qy 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGlyCys 80
   |||
Db 183 CCGCACTCGTTCACACAGAGGAGGTGTTGTATCAACAGTACGATGATGAGTGGAGGAGTGC 242
   |||

Qy 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100
   |||
Db 243 TCGCAAGCCGCCCATACATCGAACAAAGCTCAGGTAATAGCCACCAAGTTCAAGGAGAAA 302
   |||

Qy 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValThr 120
   |||
Db 303 ATCCTTGGAGTGTGTCAGCAGCCGCCCAACAAAGCTGCTCATTTGAGCCCGTAATAGCT 362
   |||

Qy 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
   |||
Db 363 TCCAACTGGCAAAAGCTTGAACCTTCTGCGCAAGCAT 401
   |||
```

RESULT 10

US-09-899-046-33
; Sequence 33, Application US/09899046
; Publication No. US2003008274A1
; GENERAL INFORMATION:

;; APPLICANT:
;; TITLE OF INVENTION: New sequences of hepatitis C virus
;; NUMBER OF SEQUENCES: 270
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/899,046
;; FILING DATE:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/362,455
;; FILING DATE:
;; INFORMATION FOR SEQ ID NO: 33:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 401 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; IMMEDIATE SOURCE:
;; CLONE: HD10-1-3
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 3..401
US-09-899-046-33

Alignment Scores:
Pred. No.: 9.85e-88 Length: 401
Score: 672.00 Matches: 124
Percent Similarity: 97.74% Conservative: 6
Best Local Similarity: 93.23% Mismatches: 3
Query Match: 96.14% Indels: 0
DB: 11 Gaps: 0

US-09-638-693-36 (1-133) x US-09-899-046-33 (1-401)

```
Qy 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
   |||
Db 3 CAAATGAAATCTGCTTGACACACCCGTCACAAAATACATTATGGCATGCTAGCT 62
   |||

Qy 21 AspleuGluValThrSerThrTrpValLeuLeuGlyValLeuAlaLeuAla 40
   |||
Db 63 GATCTGGAAGTAACACACACACCTGGGTGTTGCTTGGAGGGGTCCTCGCGCCCTAGCG 122
   |||

Qy 41 AlatyrcysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyGlyLys 60
   |||
Db 123 GCCTACTGCTGTGTCAGTCGGCTGGTGTAAATCGTGGTTCATATCGAGCTGGGGGCAAG 182
   |||

Qy 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGlyCys 80
   |||
Db 183 CCGCACTCGTTCACACAGAGGAGGTGTTGTATCAACAGTACGATGATGAGTGGAGGAGTGC 242
   |||

Qy 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100
   |||
Db 243 TCGCAAGCCGCCCATACATCGAACAAAGCTCAGGTAATAGCCACCAAGTTCAAGGAGAAA 302
   |||

Qy 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValThr 120
   |||
Db 303 ATCCTTGGAGTGTGTCAGCAGCCGCCCAACAAAGCTGCTCATTTGAGCCCGTAATAGCT 362
   |||

Qy 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
   |||
Db 363 TCCAACTGGCAAAAGCTTGAACCTTCTGCGCAAGCAT 401
   |||
```

RESULT 11

US-09-878-281-31
; Sequence 31, Application US/09878281
; Publication No. US20030032005A1
; GENERAL INFORMATION:

;; APPLICANT:
;; TITLE OF INVENTION: New sequences of hepatitis C virus
;; NUMBER OF SEQUENCES: 270
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/878,281
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/362,455
FILING DATE:
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 401 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: HD10-1-25
FEATURE:
NAME/KEY: CDS
LOCATION: 3..401
US-09-878-281-31

Alignment Scores:
Pred. No.: 9.85e-88 Length: 401
Score: 672.00 Matches: 124
Percent Similarity: 97.74% Conservative: 6
Best Local Similarity: 93.23% Mismatches: 3
Query Match: 96.14% Indels: 0
DB: 11 Gaps: 0

US-09-638-693-36 (1-133) x US-09-878-281-31 (1-401)

QY 1 GlnAsnGluLeuCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
|||||
Db 3 CAAATGAATCGTTGACACACCCCGTCACAAAATACATTATGGCATGTCAGCT 62
QY 21 AspLeuGluValThrSerThrTrpValLeuGluGlyValLeuAlaAlaLeuAla 40
|||||
Db 63 GATCTGGAAGTAGTACACACACCGCTGGGTGTTGCTTGGAGGGTCTCTCGGCCCTTAGCG 122
QY 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyLys 60
|||||
Db 123 GCCTACTGCTTGCAGTCAGCGCTGCGTTGTAATCGTGGGTGCATATCGAGCTGGGGGCAAG 182
QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGluCys 80
|||||
Db 183 CCGGCACCTCGTCCAGACAGAGGCTGTTGTATCAACAGTACGATGAGATGGAGGAGTGC 242
QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100
|||||
Db 243 TCGCAAGCGCCCATACATCGACACAGCTCAGGTAATAGCCACCAGCTTCAAGGAGAAA 302
QY 101 ValLeuGlyLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValThr 120
|||||
Db 303 ATCTTGGAGTCTGTCAGCGAGCGACCCCAACAAACAGCTGTCATTGAGCCCGCTAATAGCT 362
QY 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
|||||
Db 363 TCCAACTGGCAAAAGCTTGAACCTTCTGGCAACAGCAT 401

RESULT 12

US-09-878-281-33

Sequence 33, Application US/09878281

Publication No. US20030032005A1

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: New sequences of hepatitis C virus

TITLE OF INVENTION: genotypes for diagnosis, prophylaxis and therapy.

NUMBER OF SEQUENCES: 270

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/878,281
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/362,455
FILING DATE:
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 401 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: HD10-1-3
FEATURE:
NAME/KEY: CDS
LOCATION: 3..401
US-09-878-281-33

Alignment Scores:
Pred. No.: 9.85e-88 Length: 401
Score: 672.00 Matches: 124
Percent Similarity: 97.74% Conservative: 6
Best Local Similarity: 93.23% Mismatches: 3
Query Match: 96.14% Indels: 0
DB: 11 Gaps: 0

US-09-638-693-36 (1-133) x US-09-878-281-33 (1-401)

QY 1 GlnAsnGluLeuCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
|||||
Db 3 CAAATGAATCGTTGACACACCCCGTCACAAAATACATTATGGCATGTCAGCT 62
QY 21 AspLeuGluValThrSerThrTrpValLeuGluGlyValLeuAlaAlaLeuAla 40
|||||
Db 63 GATCTGGAAGTAGTACACACACCGCTGGGTGTTGCTTGGAGGGTCTCTCGGCCCTTAGCG 122
QY 41 AlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyLys 60
|||||
Db 123 GCCTACTGCTTGCAGTCAGCGCTGCGTTGTAATCGTGGGTGCATATCGAGCTGGGGGCAAG 182
QY 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGluCys 80
|||||
Db 183 CCGGCACCTCGTCCAGACAGAGGCTGTTGTATCAACAGTACGATGAGATGGAGGAGTGC 242
QY 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100
|||||
Db 243 TCGCAAGCGCCCATACATCGACACAGCTCAGGTAATAGCCACCAGCTTCAAGGAGAAA 302
QY 101 ValLeuGlyLeuGlnArgAlaThrGlnGlnGlnAlaValIleGluProIleValThr 120
|||||
Db 303 ATCTTGGAGTCTGTCAGCGAGCGACCCCAACAAACAGCTGTCATTGAGCCCGCTAATAGCT 362
QY 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
|||||
Db 363 TCCAACTGGCAAAAGCTTGAACCTTCTGGCAACAGCAT 401

RESULT 13

US-09-742-659-3

Sequence 3, Application US/09742659

Patent No. US20010034019A1

GENERAL INFORMATION:

APPLICANT: Hong, Zhi

APPLICANT: Butkiewicz, Nancy J.

APPLICANT: Zhong, Weidong

APPLICANT: Ingravallo, Paul

APPLICANT: Wright-Minogue, Jacquelyn

APPLICANT: Lau, Johnson Y.

APPLICANT: Lemon, Stanley M.
TITLE OF INVENTION: Chimeric HCV/GBV-B viruses
FILE REFERENCE: ID01116
CURRENT APPLICATION NUMBER: US/09/742,659
CURRENT FILING DATE: 2000-12-21
PRIOR APPLICATION NUMBER: US 60/171,469
PRIOR FILING DATE: 1999-12-22
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 9646
TYPE: DNA
ORGANISM: Hepatitis C virus
US-09-742-659-3

Alignment Scores:
Pred. No.: 1.67e-66 Length: 9646
Score: 539.00 Matches: 101
Percent Similarity: 83.46% Conservative: 10
Best Local Similarity: 75.94% Mismatches: 22
Query Match: 77.11% Indels: 0
DB: 9 Gaps: 0

US-09-638-693-36 (1-133) x US-09-742-659-3 (1-9646)

Qy 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
Db 5235 CAGATGAGTCAACCTGACGACCCCAATACCAAAATACATCATGATGCTCGGCC 5294
Qy 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyValLeuAlaLeuAla 40
Db 5295 GACCTGGAGTTCGTCACGAGCAGCCCTGGTGTCTTGGCGGCTGCTGCTCGGCC 5354
Qy 41 AlaTrpCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyLys 60
Db 5355 GCGTATTCCTGTCACAGGCTGCGTGTATGATGGCGAGGATGCTTGTTCGGGAAG 5414
Qy 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGluCys 80
Db 5415 CCGCAATTATACCTGACAGGAGGTTCTTACCAAGGATTCGATGATGGAAGAGTGC 5474
Qy 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100
Db 5475 TCTCAGCACTTACCGTACATCGAGCAAGGATGCTGCTGATGAGTTCAAGCAGAAG 5534
Qy 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValThr 120
Db 5535 GCCCTCGGCTCTCGCAGACCGGCTCCGCCAAGCAGAGGTTATCACCCCTGCTGTCCAG 5594
Qy 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
Db 5595 ACCAACTGGCAGAACTCGAGGTTCTTCTGGCGGAAGCAC 5633

RESULT 14

US-09-238-076-1
Sequence 1, Application US/09238076
Patent No. US20020102540A1
GENERAL INFORMATION:
APPLICANT: RICE, CHARLES et al.
TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
TITLE OF INVENTION: VIRUS (HCV) AND USES THEREOF
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOWELL & HAFERKAMP, L.C.
STREET: 7733 FORSYTH BLVD., SUITE 1400
CITY: ST. LOUIS
STATE: MO
COUNTRY: USA
ZIP: 63105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/238,076
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/034,756
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: HOLLAND, DONALD R.
REGISTRATION NUMBER: 35,197
REFERENCE/DOCKET NUMBER: 6029-4831
TELECOMMUNICATION INFORMATION:
TELEPHONE: 314-727-5188
TELEFAX: 314-727-6092
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9646 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-09-238-076-1
Alignment Scores:
Pred. No.: 1.67e-66 Length: 9646
Score: 539.00 Matches: 101
Percent Similarity: 83.46% Conservative: 10
Best Local Similarity: 75.94% Mismatches: 22
Query Match: 77.11% Indels: 0
DB: 9 Gaps: 0

US-09-638-693-36 (1-133) x US-09-238-076-1 (1-9646)

Qy 1 GlnAsnGluIleCysLeuThrHisProIleThrLysTyrIleMetAlaCysMetSerAla 20
Db 5235 CAGATGAGTCAACCTGACGACCCCAATACCAAAATACATCATGATGCTCGGCC 5294
Qy 21 AspLeuGluValThrSerThrTrpValLeuLeuGlyValLeuAlaLeuAla 40
Db 5295 GACCTGGAGTTCGTCACGAGCAGCCCTGGTGTCTTGGCGGCTGCTGCTCGGCC 5354
Qy 41 AlaTrpCysLeuSerValGlyCysValValIleValGlyHisIleGluLeuGlyLys 60
Db 5355 GCGTATTCCTGTCACAGGCTGCGTGTATGATGGCGAGGATGCTTGTTCGGGAAG 5414
Qy 61 ProAlaIleValProAspLysGluValLeuTyrGlnGlnTyrAspGluMetGluGluCys 80
Db 5415 CCGCAATTATACCTGACAGGAGGTTCTTACCAAGGATTCGATGATGGAAGAGTGC 5474
Qy 81 SerGlnAlaAlaProTyrIleGluGlnAlaGlnValIleAlaHisGlnPheLysGlyLys 100
Db 5475 TCTCAGCACTTACCGTACATCGAGCAAGGATGCTGCTGATGAGTTCAAGCAGAAG 5534
Qy 101 ValLeuGlyLeuLeuGlnArgAlaThrGlnGlnAlaValIleGluProIleValThr 120
Db 5535 GCCCTCGGCTCTCGCAGACCGGCTCCGCCAAGCAGAGGTTATCACCCCTGCTGTCCAG 5594
Qy 121 ThrAsnTrpGlnLysLeuGluAlaPheTrpHisLysHis 133
Db 5595 ACCAACTGGCAGAACTCGAGGTTCTTCTGGCGGAAGCAC 5633

RESULT 15

US-09-995-937-1
Sequence 1, Application US/0995937
Publication No. US20030028010A1
GENERAL INFORMATION:
APPLICANT: RICE, CHARLES et al.
TITLE OF INVENTION: FUNCTIONAL DNA CLONE FOR HEPATITIS C
TITLE OF INVENTION: VIRUS (HCV) AND USES THEREOF
NUMBER OF SEQUENCES: 21

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 11:16:00 ; Search time 27 Seconds
(without alignments)
473.720 Million cell updates/sec

Title: US-09-638-693-36
Perfect score: 699
Sequence: 1 QNEICLTHPTITRYINACMSA.....VIEPIVTNNQKLEAFWHKH 133

Scoring table: BLOSUM62.
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	691	98.9	209	2	PC1306
2	672	96.1	142	2	PC1307
3	534	76.4	3011	1	GNWVC3
4	534	76.4	3011	1	GNWVC3
5	532	76.1	492	2	PS0326
6	529	75.7	716	2	JQ1366
7	529	75.7	3011	1	S40770
8	494	70.7	3010	1	S18030
9	493	70.5	3010	1	GNWVCJ
10	492	70.4	3010	1	GNWVC
11	491	70.2	3010	1	A45573
12	490	70.1	3010	1	GNWVTW
13	484	69.2	3014	1	JC5620
14	424	60.7	876	2	PC2219
15	423	60.5	3033	1	JQ1303
16	422	60.4	3033	1	GNWVJ8
17	384	54.9	125	2	S35629
18	132	18.9	41	2	PQ0560
19	131	18.7	41	2	PQ0564
20	131	18.7	41	2	PQ0562
21	131	18.7	41	2	PQ0563
22	131	18.7	41	2	PQ0565
23	113	16.2	41	2	PQ0561
24	83	11.9	1108	2	AF1047
25	79.5	11.4	343	2	S75435
26	77	11.0	1107	2	B91271
27	77	11.0	1107	2	B86112
28	77	11.0	1107	2	E65226
29	74.5	10.7	317	2	T33277

30 74.5 10.7 346 2 B97014
31 74 10.6 370 2 S49583
32 73 10.4 221 2 S62958
33 72.5 10.4 163 2 G71463
34 72.5 10.4 546 2 T08210
35 71.5 10.2 456 2 A31857
36 71 10.2 427 2 JC4623
37 71 10.2 617 2 F75484
38 71 10.2 1010 1 AJHUPR
39 70.5 10.1 175 2 T51765
40 70 10.0 2761 2 T21064
41 69.5 9.9 1428 2 S62419
42 69 9.9 295 2 JN0888
43 69 9.9 451 1 F64155
44 69 9.9 674 2 AC1913
45 69 1016 2 T25433

ALIGNMENTS

RESULT 1
PC1306
genome polyprotein NS4a epitope containing region (isolate BR36-20) - hepatitis C vi.
C:Species: hepatitis C virus
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 17-Nov-2000
C:Accession: PC1306
R:Stuyver, L.; Van Arnhem, W.; Wyseur, A.; DeLeys, R.; Maertens, G.
Biochem. Biophys. Res. Commun. 192, 635-641, 1993
A:Title: Analysis of the putative E1 envelope and NS4a epitope regions of HCV type 3
A:Reference number: PC1300; MUID:93249436; PMID:7683463
A:Accession: PC1306
A:Molecule type: mRNA
A:Residues: 1-209 <STU>
A:Cross-references: DDBJ:DJ14600; NID:g303584; PIDN:BAA03449.1; PID:g303585
A:Experimental source: blood
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: nonstructural protein; polyprotein

Query Match 98.9%; Score 691; DB 2; Length 209;
Best Local Similarity 99.2%; Pred. No. 7.4e-62;
Matches 132; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QNEICLTHPTITRYINACMSADLEVTSTWVLGGVLAALAAAYCLSVGVVIGHIELGGK 60
DB 77 QNEICLTHPTITRYINACMSADLEVTSTWVLGGVLAALAAAYCLSVGVVIGHIELGGK 136
QY 61 PAIVDPKEVLYQOYDEMECSQAAPYIEQAQVIAHQFGKVLGLLQQRATQQQAVIEPIVT 120
DB 137 PAIVDPKEVLYQOYDEMECSQAAPYIEQAQVIAHQFGKVLGLLQQRATQQQAVIEPIVT 196
QY 121 TNWQKLEAFWHKH 133
DB 197 TNWQKLEAFWHKH 209

RESULT 2
PC1307
genome polyprotein NS4a epitope containing region (isolate HD10-1) - hepatitis C vir
C:Species: hepatitis C virus
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 17-Nov-2000
C:Accession: PC1307
R:Stuyver, L.; Van Arnhem, W.; Wyseur, A.; DeLeys, R.; Maertens, G.
Biochem. Biophys. Res. Commun. 192, 635-641, 1993
A:Title: Analysis of the putative E1 envelope and NS4a epitope regions of HCV type 3
A:Reference number: PC1300; MUID:93249436; PMID:7683463
A:Accession: PC1307
A:Molecule type: mRNA
A:Residues: 1-142 <STU>
A:Cross-references: DDBJ:DJ14602
A:Experimental source: blood
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: polyprotein

A:Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBIPI:121748)
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,417,423,448,532,540,556,576,623,645/Binding site: carbohydrate (AS
Query Match 70.7%; Score 494; DB 1; Length 3010;
Best Local Similarity 69.2%; Pred. No. 8.4e-41;
Matches 92; Conservative 15; Mismatches 26; Indels 0; Gaps 0;
QY 1 QNEICLTHPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSVGCVWIVGHIELGK 60
Db 1632 QNEVTLTHPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLTGTGSVIVGRILSGR 1691
QY 61 PAIYPDKKVLQYQDEMEECQAAPYIEQAQVIAHQFKGKVLGLLQRTAQQAQVIEPIVT 120
Db 1692 PAIYPDKKVLQYQDEMEECQAAPYIEQAQVIAHQFKGKVLGLLQRTAQQAQVIEPIVT 120
QY 121 TNWOKLEAFWAKH 133
Db 1752 SKWALEAFWAKH 1764
RESULT 9
GNWVCJ
genome polyprotein - hepatitis C virus (strain J)
N:Contains: capsid protein C; envelope protein M; major envelope protein E; nonstructura
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001
C:Accession: A39253; PS0086
R:Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimot
Proc. Natl. Acad. Sci. U.S.A. 87, 9524-9528, 1990
A:Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients
A:Reference number: A39253; MUID:91088550; PMID:2175903
A:Accession: A39253
A:Molecule type: genomic RNA
A:Residues: 1-3010 <XAT>
A:Cross-references: GB:D90208; NID:9221610; PIDN:BAAL4233.1; PID:g221611
R:Kato, N.; Ohkoshi, S.; Shimotohno, K.
Proc. Jpn. Acad. 65B, 219-223, 1989
A:Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari
A:Reference number: PS0085
A:Accession: PS0086
A:Molecule type: genomic RNA
A:Residues: 2650-2707 <XAT>
A:Experimental source: Japanese isolate
C:Comment: The cleavage sites of this polyprotein have not been determined.
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

F:196,209,234,250,305,325,417,423,430,448,532,556,576,623,645,1213,1255,2041,2077,224
Query Match 70.5%; Score 493; DB 1; Length 3010;
Best Local Similarity 68.4%; Pred. No. 1.1e-40;
Matches 91; Conservative 15; Mismatches 27; Indels 0; Gaps 0;
QY 1 QNEICLTHPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSVGCVWIVGHIELGK 60
Db 1632 QNEVTLTHPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLTGTGSVIVGRILSGR 1691
QY 61 PAIYPDKKVLQYQDEMEECQAAPYIEQAQVIAHQFKGKVLGLLQRTAQQAQVIEPIVT 120
Db 1692 PAIYPDKKVLQYQDEMEECQAAPYIEQAQVIAHQFKGKVLGLLQRTAQQAQVIEPIVT 120
QY 121 TNWOKLEAFWAKH 133
Db 1752 SKWALEAFWAKH 1764
RESULT 10
GNWVCJ
genome polyprotein - hepatitis C virus
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain J)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: hepatitis C virus
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 19-Jan-2001
C:Accession: A38465
R:Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E
J. Virol. 65, 1105-1113, 1991
A:Title: Structure and organization of the hepatitis C virus genome isolated from hum
A:Reference number: A38465; MUID:91140698; PMID:1847440
A:Accession: A38465
A:Molecule type: genomic RNA
A:Residues: 1-3010 <XAT>
A:Cross-references: EMBL:M58135; NID:Q329770; PIDN:AAA72945.1; PID:g329771
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstruct
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <MEE>
F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
F:1007-1615/Product: hepatitis C virus genome polyprotein
F:1230-1237/Region: nucleotide-binding motif A (P-loop)
F:1312-1317/Region: nucleotide-binding motif B
F:1316-1319/Region: DEXH motif
F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
F:196,209,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077
Query Match 70.4%; Score 492; DB 1; Length 3010;
Best Local Similarity 69.2%; Pred. No. 1.3e-40;
Matches 92; Conservative 14; Mismatches 27; Indels 0; Gaps 0;
QY 1 QNEICLTHPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSVGCVWIVGHIELGK 60
Db 1632 QNEVTLTHPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLTGTGSVIVGRILSGR 1691
QY 61 PAIYPDKKVLQYQDEMEECQAAPYIEQAQVIAHQFKGKVLGLLQRTAQQAQVIEPIVT 120
Db 1692 PAIYPDKKVLQYQDEMEECQAAPYIEQAQVIAHQFKGKVLGLLQRTAQQAQVIEPIVT 120
QY 121 TNWOKLEAFWAKH 133
Db 1752 SKWALEAFWAKH 1764
RESULT 11
A45573
genome polyprotein - hepatitis C virus (strain JT)
N:Contains: capsid protein C; envelope protein M; hepatitis C virus (strain JT)
protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C:Species: hepatitis C virus
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
 C:Accession: A45573
 R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata, Virus Res. 23, 39-53, 1992
 A:Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: S
 A:Reference number: A45573; MUID:92295714; PMID:1318627
 A:Accession: A45573
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-3010 <TAN>
 A:Cross-references: GB:D11168; GB:D01171; MID:g221612; PIDN:BAA01943.1; PID:g221613
 A:Experimental source: HCV-JT
 A>Note: sequence extracted from NCBI backbone (NCBI:106206, NCBIP:106207)
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; p-loop; polyprotein; serin
 F:2-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEE>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: nonstructural protein NS2 #status predicted <NS3>
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B
 F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>

Query Match 70.2%; Score 491; DB 1; Length 3010;
 Best Local Similarity 69.2%; Pred. No. 1.7e-40;
 Matches 92; Conservative 15; Mismatches 26; Indels 0; Gaps 0;

QY 1 QNEICLTHTPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSGVGVVGHIELGCK 60
 Db 1632 QNEITLTHPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLTGTGSGVIVGRIILSGR 1691

QY 61 PAIVDPKEVLYQOYDEMECSQAAPYIEQAQVIAHOFKGVGLGLLQATQQQAVIEPIVT 120
 Db 1692 PAVVPDREVLYREFDEMECSASHLPYIEQGMQLAEQFKALGLLQATKQAEAPVVE 1751

QY 121 TNWQKLEAFW 133
 Db 1752 SRWRALEAFWAKH 1764

RESULT 12
 GNWTVW
 genome polyprotein - hepatitis C virus (strain Taiwan)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
 Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 A>Note: host Homo sapiens (man)
 C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001
 C:Accession: A40244
 R:Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.
 Virology 188, 102-113, 1992
 A:Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the
 A:Reference number: A40244; MUID:92230206; PMID:1314449
 A:Accession: A40244
 A:Molecule type: genomic RNA
 A:Residues: 1-3010 <CHE>
 A:Cross-references: GB:M84754
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructura
 F:1-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEE>
 F:390-729/Product: nonstructural protein NS1 #status predicted <NS1>
 F:730-1006/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1007-1615/Product: nonstructural protein NS2 #status predicted <NS3>
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)
 F:1312-1317/Region: nucleotide-binding motif B

F:1316-1319/Region: DEXH motif
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>
 F:196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,
 Query Match 70.1%; Score 490; DB 1; Length 3010;
 Best Local Similarity 70.8%; Pred. No. 2.1e-40;
 Matches 92; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

QY 1 QNEICLTHTPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSGVGVVGHIELGCK 60
 Db 1632 QNEITLTHPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLTGTGSGVIVGRIILSGK 1691

QY 61 PAIVDPKEVLYQOYDEMECSQAAPYIEQAQVIAHOFKGVGLGLLQATQQQAVIEPIVT 120
 Db 1692 PAVVPDREVLYREFDEMECSASHLPYIEQGMQLAEQFKALGLLQATKQAEAPVVE 1751

QY 121 TNWQKLEAFW 130
 Db 1752 SKWRTLEAFW 1761

RESULT 13
 JCS620
 genome polyprotein - hepatitis C virus (isolate EUH1480)
 N:Contains: capsid protein C; envelope protein M; hepatitis C virus genome polyprotein
 Protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
 C:Species: hepatitis C virus
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001
 C:Accession: JCS620
 R:Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.
 Biochem. Biophys. Res. Commun. 236, 44-49, 1997
 A:Title: The complete coding sequence of hepatitis C virus genotype 5a, the predomina
 A:Reference number: JCS620; MUID:97366593; PMID:9223423
 A:Accession: JCS620
 A:Molecule type: mRNA
 A:Residues: 1-3014 <CHA>
 A:Cross-references: GB:Y13184
 A:Experimental source: genotype 5a, which predominates in South Africa
 C:Note: the translation of the nucleotide sequence is not complete in this paper
 C:Superfamily: hepatitis C virus genome polyprotein
 C:Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; p-loop; polyprotein; s
 F:2-115/Product: capsid protein C #status predicted <CPC>
 F:116-191/Product: envelope protein M #status predicted <EPM>
 F:192-389/Product: major envelope protein E #status predicted <MEE>
 F:384-408/Region: hypervariable #status predicted
 F:390-730/Product: nonstructural protein NS1 #status predicted <NS1>
 F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>
 F:1008-1616/Product: hepatitis C virus genome polyprotein
 F:1231-1238/Region: nucleotide-binding motif A (P-loop)
 F:1313-1318/Region: nucleotide-binding motif B
 F:1317-1320/Region: DEXH motif
 F:1617-1863/Product: nonstructural protein NS4a #status predicted <N4A>
 F:1864-2014/Product: nonstructural protein NS4b #status predicted <N4B>
 F:2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>
 F:2210-2249/Region: interferon sensitivity determining #status predicted

Query Match 69.2%; Score 484; DB 1; Length 3014;
 Best Local Similarity 68.4%; Pred. No. 8.4e-40;
 Matches 91; Conservative 16; Mismatches 26; Indels 0; Gaps 0;

QY 1 QNEICLTHTPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSGVGVVGHIELGCK 60
 Db 1633 QNEITLTHPTIKYIMACMSADLEVTSTWVLGGVLAALAAAYCLTGTGSGVIVGRIILSGR 1692

QY 61 PAIVDPKEVLYQOYDEMECSQAAPYIEQAQVIAHOFKGVGLGLLQATQQQAVIEPIVT 120
 Db 1693 PAITPDREVLYQOYDEMECSASLPYVDEARATAGQFKKVLGLICTAGKAEATLKPAA 1752

QY 121 TNWQKLEAFW 133
 Db 1753 SWSKAEQFWAKH 1765

QJ1303 genome polyprotein - hepatitis C virus (isolate HC-J6)
N:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)
C:Species: hepatitis C virus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 17-Nov-2000
C:Accession: JQ1303
R:Okamoto, H.; Okada, S.; Sugiyama, Y.; Kurai, K.; Iizuka, H.; Machida, A.; Miyakawa, Y. J. Gen. Virol. 72, 2697-2704, 1991
A:Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human
A:Reference number: JQ1303; MUID:92044440; PMID:1658196
A:Accession: JQ1303
A:Molecule type: genomic RNA
A:Residues: 1-3033 <OKA>
A:Cross-references: GB:D00944; NID:q221650; PIDN:BAA00792.1; PID:q221651
A:Experimental source: isolate HC-J6 from a Japanese individual
C:Superfamily: hepatitis C virus genome polyprotein
C:Keywords: ATP; glycoprotein; hydrolase; p-loop; polyprotein; serine proteinase; transmembrane
F:2-115/Product: capsid protein C #status predicted <CPC>
F:116-191/Product: envelope protein M #status predicted <EPM>
F:192-389/Product: major envelope protein E #status predicted <ME>
F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>
F:734-1010/Product: nonstructural protein NS2 #status predicted <NS2>
F:1011-1619/Product: hepacivirin #status predicted <NS3>
F:1316-1321/Region: nucleotide-binding motif B
F:1320-1323/Region: DEXH motif
F:1620-1866/Product: nonstructural protein NS4a #status predicted <NS4>
F:1867-2017/Product: nonstructural protein NS4b #status predicted <NS4b>

GenCore version 5.1.6
Copyright (C) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 10:55:59 ; Search time 19 Seconds
(without alignments)
329.187 Million cell updates/sec

Title: US-09-638-693-36

Perfect score: 699

Sequence: 1 QNEICLTHPTKYIMACMSA.....VIEPIVTNNQKLEAFWHKH 133

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	534	76.4	3011	POLG_HCV1	P26664 h genome po
2	534	76.4	3011	POLG_HCVH	P27958 h genome po
3	493	70.5	3010	POLG_HCVJA	P26662 h genome po
4	492	70.4	3010	POLG_HCVBK	P26663 h genome po
5	491	70.2	3010	POLG_HCVJT	Q00269 h genome po
6	490	70.1	3010	POLG_HCVTW	P29846 h genome po
7	423	60.5	3033	POLG_HCVJ6	P26660 h genome po
8	422	60.4	3033	POLG_HCVJ8	P26661 h genome po
9	77	11.0	1107	YJEP_ECOLI	P39285 escherichia
10	73	10.4	221	NCE3_YEAST	P53615 saccharomyc
11	71.5	10.2	337	LPXD_XANAC	Q8pm15 xanthomonas
12	71.5	10.2	456	RINI_PIG	P10775 sus scrofa
13	71	10.2	1010	PUR2_HUMAN	P22102 h trifuncti
14	70.5	10.1	175	BADR_RHOPA	O07458 rhodospseudo
15	69.5	9.9	337	LPXD_XANCP	Q8paw3 xanthomonas
16	69.5	9.9	1428	YA84_SCHPO	Q09773 schizosacch
17	69	9.9	206	RHO3_SCHCO	Q9p819 rhizobact
18	69	9.9	295	NIHL_RHOCA	P08718 rhodosphe
19	69	9.9	451	SUN_HAEIN	P44788 haemophilus
20	68.5	9.8	253	ISPD_RALSO	Q8xyw3 ralstonia s
21	67	9.6	1003	PUR2_CHICK	P21872 g trifuncti
22	66.5	9.5	208	RS4_HELPJ	Q9zjt4 helicobacte
23	66.5	9.5	258	PPNK_THEMEA	Q9x255 thermotoga
24	66	9.4	428	CISV_PSEAE	P14165 pseudomonas
25	66	9.4	511	Y4UD_RHISN	P55504 rhizobium s
26	65.5	9.4	380	NUEM_BOVIN	P34943 bos taurus
27	65.5	9.4	497	GLYA_CHLTR	O84439 chlamydia t
28	65.5	9.4	1199	MFD_SYNP3	Q55750 synchocyst
29	65	9.3	425	YLK1_CAEEL	P41949 caenorhabdi
30	65	9.3	1076	YK15_SCHPO	Q9c104 schizosacch
31	64.5	9.2	277	KSGA_CHLPN	Q9z6k0 chlamydia p
32	64.5	9.2	446	EX7L_STRPN	Q97q18 streptococc
33	64.5	9.2	448	EX7L_STRR6	Q8dpm9 streptococc

34	64.5	9.2	497	1	GLYA_CHLMU	Q9pjw0 chlamydia m
35	64.5	9.2	1324	1	MSH6_ARATH	O04716 arabidopsis
36	64	9.2	612	1	ITR2_YEAST	P30606 saccharomyc
37	64	9.2	676	1	HUTU_HUMAN	Q96n76 homo sapien
38	64	9.2	1001	1	AHMS_ARATH	Q9s7j8 arabidopsis
39	64	9.2	2469	1	TEGU_HSVSA	Q01056 herpesvirus
40	63.5	9.1	187	1	RS4_METJA	P54020 methanococc
41	63.5	9.1	395	1	ARGD_CAMJE	Q9pir7 campylobact
42	63.5	9.1	468	1	DLDR_STAAM	Q59822 staphylococ
43	63.5	9.1	492	1	MEPI_YEAST	P40260 saccharomyc
44	63	9.0	162	1	ILI5_CERAE	P40221 cercopithe
45	63	9.0	162	1	ILI5_MACMU	P48092 macaca mula

ALIGNMENTS

RESULT 1
POLG_HCV1
ID POLG_HCV1 STANDARD; PRT; 3011 AA.
AC P26664;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polypeptide [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate 1) (HCV).
DE Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11104;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9117286; PubMed=1848704;
RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
RA Bradley D.W., Kuo G., Houghton M.;
RL "Genetic organization and diversity of the hepatitis C virus.";
CC Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polypeptide, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC (RNA)(N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M62321; AAA45676.1; -.
CC PIR; A39166; GNWV3.
CC PDB; 1A1V; 16-FEB-99.
CC PDB; 1HEI; 25-NOV-98.
CC MEROPS; S29.001; -.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR002522; HCV_capsid.

DR InterPro: IPR002521; HCV_core.
 DR InterPro: IPR002519; HCV_env.
 DR InterPro: IPR002531; HCV_NS1.
 DR InterPro: IPR002518; HCV_NS2.
 DR InterPro: IPR004109; HCV_NS3.
 DR InterPro: IPR000745; HCV_NS4a.
 DR InterPro: IPR001490; HCV_NS4b.
 DR InterPro: IPR002868; HCV_NS5a.
 DR InterPro: IPR002166; HCV_RdRp.
 DR InterPro: IPR001650; Helicase.C.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR Pfam: PF01543; HCV_capsid; 1.
 DR Pfam: PF01542; HCV_core; 1.
 DR Pfam: PF01539; HCV_env; 1.
 DR Pfam: PF01560; HCV_NS1; 1.
 DR Pfam: PF01538; HCV_NS2; 1.
 DR Pfam: PF02907; HCV_NS3; 1.
 DR Pfam: PF01006; HCV_NS4a; 1.
 DR Pfam: PF01001; HCV_NS4b; 1.
 DR Pfam: PF01506; HCV_NS5a; 1.
 DR Pfam: PF00271; Helicase.C; 1.
 DR Pfam: PF00998; Viral_RdRp; 1.
 DR ProDom: PD186062; HCV_NS1; 1.
 DR SMART: SM00487; DEXDC; 1.
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 KW 3D-structure.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1616 1862 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 2014 3011 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT TRANSMEM 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
 FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT NP_BIND 1230 1237 ATP (POTENTIAL).
 FT SITE 1316 1319 DECH BOX.
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCESAF9 CRC64;
 Query Match 76.4%; Score 534; DB 1; Length 3011;
 Best Local Similarity 74.4%; Pred. No. 3,9e-45;
 Matches 99; Conservative 12; Mismatches 22; Indels 0; Gaps 0;

Qy 1 QNEICLTHPTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSGVGVIVGHIELGSK 60

Db 1632 QNEITLTHPTKYIMTCNSADLEVTSTWVLLGGVLAALAAAYCLSGVGVIVGRVLSGK 1691
 Qy 61 PAIYDPEKLYQOYDEMEECSSQAAPYIEQAQVIAHOFKGVKVLGLLQRTQQOAVLEPIVT 120
 Db 1692 PAIIPDREVLRYEFDEMEECSSQHLPIYIEGGMMLAEQFKOKALGLLQRTASROAEVITAPVQ 1751
 Qy 121 TNWQLEAFWHKH 133
 Db 1752 TNWQLETFFWAKH 1764
 RESULT 2
 POLG_HCVH STANDARD; PRT; 3011 AA.
 AC P27958;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.99.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate H) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11108;
 RP [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92052256; PubMed=1658800;
 RA Inchauspe G., Zebedee S., Lee D.H., Sugitani M., Nasoff M.,
 RA Prince A.M.;
 RT "Genomic structure of the human prototype strain H of hepatitis C
 RT virus: comparison with American and Japanese isolates.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10292-10296(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1207-1657.
 RX MEDLINE=97331322; PubMed=9187654;
 RA Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C.;
 RT "Structure of the hepatitis C virus RNA helicase domain.";
 RL Nat. Struct. Biol. 4:463-467(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.
 RX MEDLINE=98154321; PubMed=9493270;
 RA Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A.,
 RA Murcko M.A., Lin C., Caron P.R.;
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound
 RT oligonucleotide: the crystal structure provides insights into the mode
 RT of unwinding.";
 RL Structure 6:89-100(1998).
 CC -1- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.
 CC -1- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF
 CC NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.
 CC -1- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE
 CC ACTIVATION OF NS3.
 CC -1- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.
 CC -1- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN
 CC ESSENTIAL ROLE IN THE VIRUS REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the p6
 CC position. Cys or Thr in p1 and Ser or Ala in p1',
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1
 CC AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.
 CC -1- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY
 CC PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.
 CC -1- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.
 CC -1- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.


```

FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2768 2768 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT STRAND 1031 1035
FT STRAND 1039 1047
FT STRAND 1050 1050
FT STRAND 1059 1063
FT STRAND 1068 1074
FT TURN 1075 1076
FT STRAND 1077 1081
FT STRAND 1082 1085
FT TURN 1086 1087
FT STRAND 1090 1092
FT TURN 1093 1094
FT STRAND 1095 1097
FT STRAND 1101 1103
FT TURN 1104 1107
FT STRAND 1108 1112
FT STRAND 1120 1120
FT STRAND 1122 1122
FT STRAND 1129 1133
FT TURN 1135 1136
FT STRAND 1139 1144
FT STRAND 1149 1157
FT TURN 1158 1161
FT TURN 1162 1163
FT TURN 1165 1166
FT STRAND 1168 1171
FT TURN 1172 1174
FT STRAND 1175 1186
FT TURN 1187 1188
FT STRAND 1189 1197
FT TURN 1198 1202
FT TURN 1203 1204
FT STRAND 1680 1688
SQ SEQUENCE 3010 AA; 327189 MW; F8422D5ECCDFD9C CRC64;

Query Match 70.4%; Score 492; DB 1; Length 3010;
Best Local Similarity 69.2%; Pred. No. 6.2e-41;
Matches 92; Conservative 14; Mismatches 27; Indels 0; Gaps 0;

Qy 1 QNEICLTHPIKYMACHSADLEVTTSWVLGGVLAALAAAYCLSVGCVIIVGHIELGGK 60
Db 1632 QNEVTLTHPIKYMACHSADLEVTTSWVLGGVLAALAAAYCLTTSVIVGRIISGR 1691
Qy 61 PAIVDPKEVLYQQYDEMEECQAAPYIEQAQVIAHQFKVGLGLQLRATQQAQVIEPIVT 120
Db 1692 PAIVDPRELLYQDEMEECASHLPYIEQGMQLAEQFKQKALGLQLQTATKQAEAAAPWE 1751
Qy 121 TNWQKLEAFWHKH 133
Db 1752 SKWRALETFWAKH 1764

RESULT 5
POLG_HCVJT STANDARD; PRT; 3010 AA.
AC Q00269;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2

```

```

DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-JT) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=31642;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295714; PubMed=1318627;
RA Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J.,
RA Nakazawa T., Hijikata M., Ishimura Y., Shimotohno K.;
RT "Molecular cloning of hepatitis C virus genome from a single Japanese
RT carrier: sequence variation within the same individual and among
RT infected individuals.";
RL Virus Res. 23:39-53(1992).
CC -|- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC NS3 AND NS5 MAY PLAY A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC -|- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -|- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC {RNA}(N).
CC -|- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -|- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D11168; BAA01943.1; -
DR PIR; A45573; A45573.
DR PDB; 1AIQ; 25-MAR-98.
DR PDB; 1JXP; 14-JAN-98.
DR MEROPS; S29.001; -
DR MEROPS; U39.001; -
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR007095; RNA_pol_DS_Ps.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXdc; 1.
DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW

```

Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 3D-structure.

INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.

CHAIN 1 115
 CHAIN 116 191
 CHAIN 192 383
 CHAIN 384 729
 CHAIN 730 1006
 CHAIN 1007 1635
 CHAIN 1636 1862
 CHAIN 1863 2013
 CHAIN 2014 3010
 CHAIN 3011 369
 CHAIN 370 1083
 CHAIN 1084 1107
 CHAIN 1108 1165
 CHAIN 1166 1237
 CHAIN 1238 1316
 CHAIN 1317 196
 CHAIN 197 209
 CHAIN 210 234
 CHAIN 235 250
 CHAIN 251 305
 CHAIN 306 417
 CHAIN 418 423
 CHAIN 424 430
 CHAIN 431 448
 CHAIN 449 532
 CHAIN 533 540
 CHAIN 541 556
 CHAIN 557 576
 CHAIN 577 623
 CHAIN 624 645
 CHAIN 646 2041
 CHAIN 2042 2077
 CHAIN 2078 2240
 CHAIN 2241 2529
 CHAIN 2530 2788
 CHAIN 2789 3010

AA; 326573 MW; 94A1C77435D642BB CRC64;
 Query Match 70.2%; Score 491; DB 1; Length 3010;
 Best Local Similarity 69.2%; Pred. No. 7.9e-41;
 Matches 92; Conservative 15; Mismatches 26; Indels 0; Gaps 0;

QY 1 ONEICLTHTTKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSGVGVVGHIELGK 60
 DB 1632 ONEITHTPTKFMACMSADLEVTSTWVLGGVLAALAAAYCLTGTGTVVIGRIILSGR 1691
 QY 61 PAIVDPKEVLYQOYDEMEECSSQAAPYIEQAQVIAHOFKGVLLGLQRATQQAQVIEPIVT 120
 DB 1692 PAVVPDREVLRYFEDMEECASHLPYIEQGMQLAEQFKQKGLLGLTATKQEAAPVVE 1751
 QY 121 TNNQKLEAFWKH 133
 DB 1752 SRWRALEAFWAKH 1764

RESULT 6
 POLG_HCVTW STANDARD; PRT; 3010 AA.
 AC P29846;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Capsid protein c (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

OS Hepatitis C virus (isolate Taiwan) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=31645;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92230206; PubMed=1314449;
 RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;
 RT "The Taiwanese hepatitis C virus genome: sequence determination and
 mapping the 5' termini of viral genomic and antigenomic RNA";
 RL Virology 188:102-113(1992).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
 CC (RNA)(N).
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: M84754; -; NOT_ANNOTATED_CDS.
 CC PIR: A40244; GHWVTV.
 CC PDB: 1NS4; 25-FEB-03.
 CC PDB: 1NS3; 08-APR-98.
 CC MEROPS: S29.001; -.
 CC MEROPS: U39.001; -.
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_env.
 CC InterPro: IPR002531; HCV_NS1.
 CC InterPro: IPR002518; HCV_NS2.
 CC InterPro: IPR004109; HCV_NS3.
 CC InterPro: IPR000745; HCV_NS4a.
 CC InterPro: IPR001490; HCV_NS4b.
 CC InterPro: IPR002868; HCV_NS5a.
 CC InterPro: IPR002166; HCV_RdRP.
 CC InterPro: IPR007095; RNA_pol_DS_PS.
 CC InterPro: IPR007094; RNA_pol_PSVir.
 CC Pfam: PF01543; HCV_capsid; 1.
 CC Pfam: PF01542; HCV_core; 1.
 CC Pfam: PF01539; HCV_env; 1.
 CC Pfam: PF01560; HCV_NS1; 1.
 CC Pfam: PF01538; HCV_NS2; 1.
 CC Pfam: PF02907; HCV_NS3; 1.
 CC Pfam: PF01006; HCV_NS4a; 1.
 CC Pfam: PF01001; HCV_NS4b; 1.
 CC Pfam: PF01506; HCV_NS5a; 1.
 CC Pfam: PF00271; helicase_C; 1.
 CC Pfam: PF00998; Viral_RdRP; 1.
 CC ProDom: PD186062; HCV_NS1; 1.
 CC SMART: SM00487; DEXDc; 1.
 CC PolyProtein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 CC Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
 CC Transmembrane; Nonstructural protein; Hydrolase; Serine protease;
 CC 3D-structure.

INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 CHAIN 1 115
 CHAIN 116 191
 CHAIN 192 383
 CHAIN 384 729
 CHAIN 730 1006
 CHAIN 1007 1635
 CHAIN 1636 1862
 CHAIN 1863 2013
 CHAIN 2014 3010
 CHAIN 3011 369
 CHAIN 370 1083
 CHAIN 1084 1107
 CHAIN 1108 1165
 CHAIN 1166 1237
 CHAIN 1238 1316
 CHAIN 1317 196
 CHAIN 197 209
 CHAIN 210 234
 CHAIN 235 250
 CHAIN 251 305
 CHAIN 306 417
 CHAIN 418 423
 CHAIN 424 430
 CHAIN 431 448
 CHAIN 449 532
 CHAIN 533 540
 CHAIN 541 556
 CHAIN 557 576
 CHAIN 577 623
 CHAIN 624 645
 CHAIN 646 2041
 CHAIN 2042 2077
 CHAIN 2078 2240
 CHAIN 2241 2529
 CHAIN 2530 2788
 CHAIN 2789 3010

```

FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1616 1862 PROTEASE/HELICASE NS4 (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 DECH BOX.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 233 233 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3010 AA; 327047 MW; AAD267D55CDFE215 CRC64;

Query Match 70.1%; Score 490; DB 1; Length 3010;
Best Local Similarity 70.8%; Pred. No. 9.9e-41;
Matches 92; Conservative 13; Mismatches 25; Indels 0; Gaps 0;

QY 1 ONEICLTHPIKTYIMACMSADLEVTSTWVLGGVLAALAAAYCLSVGCVVIGHIELGK 60
Db 1632 ONEVLTHTPIKTYIMACMSADLEVTSTWVLGGVLAALAAAYCLTGTSGVIVGRIILSGK 1691
QY 61 PAIYDPKVELYQXDEMECECSQAAPYIEQAQVIAHQFKGKVLGLLQRAOQOAVIEPIVT 120
Db 1692 PAVYDPREVLYQXDEMECECSHLPYIEQGHQLAEQFKQKALGLLQTAIKQAEAAAPVE 1751
QY 121 TSNQKLEAFW 130
Db 1752 SKWRTLEAFW 1761

RESULT 7
POLG_HCVJ6 STANDARD; PRT; 3033 AA.
AC P28660;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66); Nonstructural protein NS5A (P56); Nonstructural protein
OS Hepatitis C virus (isolate HC-J6) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_Taxid-11113;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=92044440; PubMed=1658196;
RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,
RA Machida A., Miyakawa Y., Mayumi M.;
RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated
RT from a human carrier: comparison with reported isolates for conserved
RT and divergent regions.";
RL J. Gen. Virol. 72:2697-2704(1991).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polyprotein, commonly with Asp or Glu in the P6
CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC [RNA](N).
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D00944; BAA00792.1; -
CC PIR; JQ1303; JQ1303.
CC HSSP; P27958; 1HEI.
CC MEROPS; S29.001; -
CC MEROPS; U39.001; -
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR002522; HCV_capsid.
CC InterPro; IPR002521; HCV_core.
CC InterPro; IPR002519; HCV_env.
CC InterPro; IPR002531; HCV_NS1.
CC InterPro; IPR002518; HCV_NS2.
CC InterPro; IPR004109; HCV_NS3.
CC InterPro; IPR000745; HCV_NS4a.
CC InterPro; IPR001490; HCV_NS4b.
CC InterPro; IPR002868; HCV_NS5a.
CC InterPro; IPR002166; HCV_RdRp.
CC InterPro; IPR001650; Helicase_C.
CC InterPro; IPR007095; RNA_pol_PSv.
CC InterPro; IPR007094; RNA_pol_Psvir.
CC Pfam; PF01543; HCV_capsid; 1.
CC Pfam; PF01542; HCV_core; 1.
CC Pfam; PF01539; HCV_env; 1.
CC Pfam; PF01560; HCV_NS1; 1.
CC Pfam; PF01538; HCV_NS2; 1.
CC Pfam; PF02907; HCV_NS3; 1.
CC Pfam; PF01006; HCV_NS4a; 1.
CC Pfam; PF01001; HCV_NS4b; 1.
CC Pfam; PF01506; HCV_NS5a; 1.
CC Pfam; PF00271; helicase_C; 1.
CC Pfam; PF00998; Viral_RdRp; 1.
CC Pfam; PF018062; HCV_NS1; 1.
CC ProDom; PD186062; HCV_NS1; 1.
CC SMART; SM00487; DEXdc; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 115 CELLULAR AMINOPEPTIDASE.
FT CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 733 NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
FT CHAIN 734 1010 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1011 1619 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1620 1866 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).

```

```

FT CHAIN 1867 2017 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2018 3033 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1087 1087 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1111 1111 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1169 1169 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1234 1241 ATP (POTENTIAL).
FT SITE 1320 1323 DECH_BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 542 542 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 627 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1091 1091 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2038 2038 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2811 2811 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BE9E CRC64;

Query Match 60.5%; Score 423; DB 1; Length 3033;
Best Local Similarity 57.6%; Pred. No. 5.1e-34;
Matches 76; Conservative 20; Mismatches 36; Indels 0; Gaps 0;

QY 2 NEICLTHPTIKYIMACMSGLDVTWVLLGGVLAALAAAYCLSVGVVIVGHIELGKPK 61
DB 1637 NEVTLTHPTVKYIATCMQADLEVTSTWVLLAGVLAANAAVYCLATGCVCIIRLHVQRA 1696
QY 62 AIVDPKEVLYOQYDEMEECQAAYIEQAQVIAHQFKGKVLGLLQARATQQAQVIEPIVTT 121
DB 1697 VVAPDKEVLYEAFDEMEECASRAALIEEQRIAEMLKSKIGLLQQAQSKAQDIQPAVQA 1756
QY 122 NQKLEAFWHKH 133
DB 1757 SMPKVEQFWAKH 1768

RESULT 8
POLG_HCVJ8
ID POLG_HCVJ8 STANDARD; PRT; 3033 AA.
AC P26661;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Contains annotation update)
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22);
DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)
DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirin)
DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
OS Hepatitis C virus (isolate HC-J8) (HCV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OC NCBI_TaxID=11115;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92230232; PubMed=1314459;
RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,
RA Fukuda S., Tsuda F., Mishiro S.;
RT "Full-length sequence of a hepatitis C virus genome having poor
RT homology to reported isolates: comparative study of four distinct
RT genotypes."
RL Virology 188:331-341(1992).
RC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE

```

```

CC CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC CC precursor polyprotein, commonly with Asp or Glu in the P6
CC CC position, Cys or Thr in P1 and Ser or Ala in P1'.
CC CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC CC [RNA](N).
CC CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC CC LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC CC PROTEIN C AND MRNA.
CC CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; D10988; BAA01761.1; -
CC CC PIR; A40250; GNVJ8.
CC CC HSP; P27938; IHEI.
CC CC MEROPS; S29.001; -
CC CC MEROPS; U39.001; -
CC CC InterPro; IPR001410; DEAD.
CC CC InterPro; IPR002522; HCV_capsid.
CC CC InterPro; IPR002521; HCV_core.
CC CC InterPro; IPR002519; HCV_env.
CC CC InterPro; IPR002531; HCV_NS1.
CC CC InterPro; IPR002518; HCV_NS2.
CC CC InterPro; IPR004109; HCV_NS3.
CC CC InterPro; IPR000745; HCV_NS4a.
CC CC InterPro; IPR001490; HCV_NS4b.
CC CC InterPro; IPR002868; HCV_NS5a.
CC CC InterPro; IPR002166; HCV_RdRP.
CC CC InterPro; IPR007095; RNA_pol_DS_PS.
CC CC InterPro; IPR007094; RNA_pol_PSVir.
CC CC Pfam; PF01543; HCV_capsid; 1.
CC CC Pfam; PF01542; HCV_core; 1.
CC CC Pfam; PF01539; HCV_env; 1.
CC CC Pfam; PF01560; HCV_NS1; 1.
CC CC Pfam; PF01538; HCV_NS2; 1.
CC CC Pfam; PF02907; HCV_NS3; 1.
CC CC Pfam; PF02907; HCV_NS4a; 1.
CC CC Pfam; PF01006; HCV_NS4a; 1.
CC CC Pfam; PF01001; HCV_NS4b; 1.
CC CC Pfam; PF01506; HCV_NS5a; 1.
CC CC Pfam; PF00998; Viral_RdRP; 1.
CC CC ProDom; PD186062; HCV_NS1; 1.
CC CC SMART; SM00487; DEXDc; 1.
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;
KW Transmembrane; Nonstructural
FT INIT_MET 1 1
FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 733 NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
FT CHAIN 734 1010 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1011 1619 PROTEASE/HELICASE NS3 (POTENTIAL).
FT CHAIN 1620 1866 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 1867 2017 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2018 3033 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1087 1087 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1111 1111 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1169 1169 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1234 1241 ATP (POTENTIAL).
FT SITE 1320 1323 DECH_BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).

```



```

RL J. Cell Biol. 133:1017-1026(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Duesterhoeft A., Floeth M., Fritz C., Heuss-Neitzel D.,
RA Hilbert H., Moestl D.;
RA Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN A NOVEL PATHWAY OF EXPORT OF PROTEINS THAT
CC LACK A CLEAVABLE SIGNAL SEQUENCE. APPEARS TO ENCODE AN ENDOGENOUS
CC NON-CLASSICAL EXPORT SUBSTRATE.
CC -!- SIMILARITY: BELONGS TO THE PLANT AND PROKARYOTIC CARBONIC
CC ANHYDRASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U52369; AAC49352.1; -
DR EMBL; 271312; CAA95901.1; -
DR PIR; S62958; S62958.
DR SGD; S0004981; NCE103.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR InterPro; IPR001765; Prok_Coanhd.
DR Pfam; PF00484; ProCA.1.
DR PROSITE; PS00704; PROK_CO2_ANHYDRASE_1; FALSE_NEG.
DR PROSITE; PS00705; PROK_CO2_ANHYDRASE_2; FALSE_NEG.
FT CONFLICT 39 L -> W (IN REF. 1).
SQ SEQUENCE 221 AA; 24859 MW; 96D0EBAABBD9790 CRC64;

Query Match 10.4%; Score 73; DB 1; Length 221;
Best Local Similarity 24.7%; Pred. No. 3.7;
Matches 24; Conservative 14; Mismatches 41; Indels 18; Gaps 4;

QY 17 CMSADLEVTSTWVLGGVLAALAAAYCLSVGCVIVGHIEGK-KPAIVDPKREVLYQQYD 75
Db 84 CHSEDLTKAT-----LEFAILCLAVNKVILGGHFDGCGIKTCLTNQREAL----- 129

QY 76 EMEECQAAPYIEQAQVIAHQKFGKVLGLLQRTAQOQ 112
Db 130 PKVNCSHLYKYLDDIDTYHE---ESQNLHLKLTQRE 163

RESULT 11
LPXD_XANAC STANDARD; PRT; 337 AA.
AC Q8PML5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acetyltransferase
DE (EC 2.3.1.-)
GN LPXD OR XAC1411.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorfy H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.T.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,

```

```

RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezra R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
CC -!- CATALYTIC ACTIVITY: UDP-3-O-(3-hydroxytetradecanoyl)glucosamine +
CC (R)-3-hydroxytetradecanoyl-[acyl-carrier protein] = UDP-2,3-bis(3-
CC hydroxytetradecanoyl)glucosamine + [acyl-carrier protein].
CC -!- PATHWAY: Lipid A biosynthesis; third step.
CC -!- SIMILARITY: BELONGS TO THE TRANSFERASE HEXAPEPTIDE REPEAT FAMILY.
CC LPXD SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE011772; AAM36282.1; -
DR HAMAP; MF_00523; -; 1.
DR InterPro; IPR001451; Hexapep_transf.
DR Pfam; PF00132; hexapep; 8.
DR Pfam; PF04613; LpxD; 1.
DR PROSITE; PS00101; HEXAPEP_TRANSFERASES; FALSE_NEG.
DR Transferase; Acyltransferase; Lipid A biosynthesis; Lipid synthesis;
KW Repeat; Complete proteome.
KW SEQUENCE 337 AA; 34960 MW; 199694075EA3ED5C CRC64;

Query Match 10.2%; Score 71.5; DB 1; Length 337;
Best Local Similarity 29.6%; Pred. No. 8.2;
Matches 21; Conservative 11; Mismatches 38; Indels 1; Gaps 1;

QY 37 AALAAVCLSGCVIVGHIEGKPAIVDPKREVLYQQYDEMEECQAAPYIEQAQVIAHQ 96
Db 258 AKIGRYCLLGGHVGVGHLIECDK-VVITGKSVVRNSIHPEGYSSTGPTLTDNRTRKNA 316

QY 97 FRGKVLGLLQRT 107
Db 317 AREKQLDVLAR 327

RESULT 12
RINI_PIG STANDARD; PRT; 456 AA.
AC P10775;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ribonuclease inhibitor.
DE RNH OR RI.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=89118268; PubMed=3219361;
RA Hofsteenge J., Kieffer B., Matthies R., Hemmings B.A., Stone S.R.;
RT "Amino acid sequence of the ribonuclease inhibitor from porcine liver
RT reveals the presence of leucine-rich repeats."
RL Biochemistry 27:8537-8544(1988).
RN [2]
RP SEQUENCE OF 82-456 FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=91104783; PubMed=2271559;
RA Vicentini A.M., Kieffer B., Matthies R., Meyhack B., Hemmings B.A.,
RA Stone S.R., Hofsteenge J.;

```


ribonucleotide synthetase) (Phosphoribosylglycinamide synthetase);
DE Phosphoribosyl-aminoglycinamide cyclo-ligase (EC 6.3.3.1) (AIRS)
DE (Phosphoribosyl-aminimidazole synthetase) (AIR synthetase);
DE Phosphoribosylglycinamide formyltransferase (EC 2.1.2.2) (GART) (GAR
DE transformylase) (5'-phosphoribosylglycinamide transformylase)].
GN GART OR PRGS.
OS Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=91067455; PubMed=2147474;
RA Aimi J., Qiu H., Williams J., Zalkin H., Dixon J.E.;
RT "De novo purine nucleotide biosynthesis: cloning of human and avian
RT cDNAs encoding the trifunctional glycinamide ribonucleotide
RT synthetase-aminimidazole ribonucleotide synthetase-glycinamide
RT ribonucleotide transformylase by functional complementation in E.
RT coli.";
RL Nucleic Acids Res. 18:6665-6672(1990).
RN [2]
RP SEQUENCE OF 709-1010 FROM N.A.
RX MEDLINE=90222141; PubMed=2183217;
RA Schild D., Brake A.J., Kiefer M.C., Young D., Barr P.J.;
RT "Cloning of three human multifunctional de novo purine biosynthetic
RT genes by functional complementation of yeast mutations.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:2916-2920(1990).
CC -1- CATALYTIC ACTIVITY: ATP + 5-phospho-D-ribo-sylamine + glycine = ADP
CC + phosphate + N(1)-(5-phospho-D-ribo-syl)glycinamide.
CC -1- CATALYTIC ACTIVITY: 10-formyltetrahydrofolate + N(1)-(5-phospho-D-
CC ribosyl)glycinamide = tetrahydrofolate + N(1)-(5-phospho-D-
CC ribosyl)-D-ribo-sylglycinamide.
CC -1- CATALYTIC ACTIVITY: ATP + 2-(formamido)-N(1)-(5-phospho-D-
CC ribosyl)acetamide = ADP + phosphate + 5-amino-1-(5-phospho-D-
CC ribosyl)imidazole.
CC -1- PATHWAY: SECOND, THIRD AND FIFTH STEPS IN DE NOVO PURINE
CC BIOSYNTHESIS.
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name-Long:
CC IsoId=P22102-1; Sequence-Displayed;
CC Name-Short:
CC IsoId=P22102-2; Sequence-VSP_005517;
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE GARS FAMILY.
CC -1- SIMILARITY: TO OTHER AIRS AND GART FROM BACTERIA AND EUKARYOTES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X54199; CAA381119.1; -;
DR EMBL; M32082; AAA60077.1; -;
DR PIR; S12616; AJHUPR.
DR HSSP; P15640; IGSO.
DR Genew; HGNC:4163; GART.
DR MIM; 138440; -;
DR GO; GO:0004644; F:phosphoribosylglycinamide formyltransferase. . .; TAS.
DR InterPro; IPR000728; AIRS related.
DR InterPro; IPR002376; formyl_transf.
DR InterPro; IPR001115; Gars.
DR InterPro; IPR001555; GART.
DR InterPro; IPR004733; PurM_cligase.
DR InterPro; IPR004607; PurN.
DR Pfam; PF00586; AIRS; 1.
DR Pfam; PF02769; AIRS_C; 1.
DR Pfam; PF00551; formyl_transf; 1.
DR Pfam; PF01071; GARS; 1.
DR Pfam; PF02842; GARS_B; 1.
DR Pfam; PF02843; GARS_C; 1.

Pfam; PF02844; GARS_N; 1.
DR TIGRFAMS; TIGR00877; purD; 1.
DR TIGRFAMS; TIGR00878; purM; 1.
DR TIGRFAMS; TIGR00639; purN; 1.
DR PROSITE; PS00184; GARS; 1.
DR PROSITE; PS00373; GART; 1.
KW Multifunctional enzyme; Purine biosynthesis; Ligase; Transferase;
KW Alternative splicing; Polymorphism.
FT DOMAIN 1 433
FT GARS.
FT DOMAIN 434 809
FT AIRS.
FT DOMAIN 810 1010
FT GART.
FT ACT_SITE 951 951
FT BY SIMILARITY.
FT VARSPLIC 434 1010
FT Missing (in isoform Short).
FT L -> F (IN dBSNP:1804387).
FT /FTID=VAR_011817.
FT V -> I (IN dBSNP:8788).
FT /FTID=VAR_011818.
FT D -> G (IN dBSNP:8971).
FT /FTID=VAR_011819.
SQ SEQUENCE 1010 AA; 107767 MW; 9A4213F746EB17A2 CRC64;
Query Match 10.2%; Score 71; DB 1; Length 1010;
Best Local Similarity 27.0%; Pred. No. 29;
Matches 27; Conservative 15; Mismatches 32; Indels 26; Gaps 5;
QY 16 ACMSADLEVTSTWVLGGVLAALAAAYCLSVGCVIIVGHTEIGKPAIVDPKEVLVQQYD 75
Db 539 SCGKLDLSTVEA-----VVAGIAKACGKAGCAL-----LGETAEMPD---MYPP-G 581
QY 76 EMEECQAAPYIEQAQVIAHQKQKVLGLLQRAQQQAVI 115
Db 582 EYDLAGFAVGAMERDQKLPH-----LERITEGDVVV 612
RESULT 14
BADR_RHOPA
ID BADR_RHOPA STANDARD; PRT; 175 AA.
AC 007458;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcriptional activatory protein badr (Benzoate anaerobic
DE degradation regulator).
GN BADR
OS Rhodopseudomonas palustris.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Rhodopseudomonas.
OX NCBI_TaxID=1076;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CGA009;
RX MEDLINE=97322399; PubMed=9177244;
RA England P.G., Pelletier D.A., Dispensa M., Gibson J., Harwood C.S.;
RT "A cluster of bacterial genes for anaerobic benzene ring
RT biodegradation.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:6484-6489(1997).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=99194718; PubMed=10094687;
RA England P.G., Harwood C.S.;
RT "Badr, a new MarR family member, regulates anaerobic benzoate
RT degradation by Rhodopseudomonas palustris in concert with Aadr, an
RT Fnr family member.";
RL J. Bacteriol. 181:2102-2109(1999).
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR OF GENES FOR THE ANAEROBIC
CC DEGRADATION OF BENZOATE.
CC -1- SIMILARITY: BELONGS TO THE MARR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 29, 2003, 08:13:24 ; Search time 53 Seconds
(without alignments)
647.566 Million cell updates/sec

Title: US-09-638-693-36

Perfect score: 699

Sequence: 1 QNEICLTHPTIKYIMACMSA.....VIEPIVTNNQKLEAFWKKH 133

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	691	98.9	209	12 Q81594	Q81594 hepatitis c
2	675	96.6	3021	12 Q92933	Q92933 hepatitis c
3	674	96.4	3021	12 Q68870	Q68870 hepatitis c
4	673	96.3	3021	12 Q81258	Q81258 hepatitis c
5	672	96.1	133	12 Q81595	Q81595 hepatitis c
6	666	95.3	138	12 Q68233	Q68233 hepatitis c
7	661	94.6	3021	12 Q81495	Q81495 hepatitis c
8	654	93.6	138	12 Q68241	Q68241 hepatitis c
9	654	93.6	138	12 Q68239	Q68239 hepatitis c
10	616	88.1	193	12 Q56637	Q56637 hepatitis c
11	605	86.6	3023	12 Q81487	Q81487 hepatitis c
12	552	79.0	3019	12 Q68801	Q68801 hepatitis c
13	546	78.1	138	12 Q68223	Q68223 hepatitis c
14	539	77.1	3011	12 Q36579	Q36579 hepatitis c
15	538	77.0	138	12 Q68224	Q68224 hepatitis c
16	538	77.0	138	12 Q68225	Q68225 hepatitis c

17	538	77.0	3011	12 Q9IFE5	Q9IFE5 hepatitis c
18	534	76.4	2436	12 Q81756	Q81756 hepatitis c
19	534	76.4	3011	12 Q36610	Q36610 hepatitis c
20	534	76.4	3011	12 Q36609	Q36609 hepatitis c
21	534	76.4	3011	12 Q36608	Q36608 hepatitis c
22	534	76.4	3015	12 Q9PMX5	Q9PMX5 hepatitis c
23	534	76.4	3015	12 Q9PMW9	Q9PMW9 hepatitis c
24	533	76.3	138	12 Q68226	Q68226 hepatitis c
25	531	76.0	138	12 Q68236	Q68236 hepatitis c
26	530	75.8	138	12 Q68243	Q68243 hepatitis c
27	529	75.7	138	12 Q68234	Q68234 hepatitis c
28	529	75.7	313	12 Q02258	Q02258 hepatitis c
29	529	75.7	3011	12 Q03463	Q03463 hepatitis c
30	527	75.4	3011	12 Q9ELS8	Q9ELS8 hepatitis c
31	516	73.8	138	12 Q68222	Q68222 hepatitis c
32	516	73.8	3011	12 Q9DIT6	Q9DIT6 hepatitis c
33	511	73.1	3011	12 Q913D4	Q913D4 hepatitis c
34	509	72.8	3011	12 Q81754	Q81754 hepatitis c
35	505	72.2	3010	12 Q68533	Q68533 hepatitis c
36	503	72.0	3010	12 Q9J3H2	Q9J3H2 hepatitis c
37	502	71.8	3010	12 Q9J3G0	Q9J3G0 hepatitis c
38	501	71.7	138	12 Q68240	Q68240 hepatitis c
39	501	71.7	3010	12 Q93016	Q93016 hepatitis c
40	500	71.5	3008	12 Q9J3F4	Q9J3F4 hepatitis c
41	500	71.5	3010	12 Q9D7D6	Q9D7D6 hepatitis c
42	499	71.4	138	12 Q68211	Q68211 hepatitis c
43	498	71.2	2864	12 Q92973	Q92973 hepatitis c
44	498	71.2	2864	12 Q9WLK9	Q9WLK9 hepatitis c
45	498	71.2	2864	12 Q9WLK8	Q9WLK8 hepatitis c

ALIGNMENTS

RESULT 1

Q81594	ID	Q81594	PRELIMINARY;	PRT;	209 AA.
AC	Q81594;				
DT	01-NOV-1996	(Tremblrel. 01, Created)			
DT	01-NOV-1996	(Tremblrel. 01, Last sequence update)			
DT	01-MAR-2003	(Tremblrel. 23, Last annotation update)			
DE	Nonstructural protein 4 (Fragment).				
GN	NS4.				
OS	Hepatitis C virus.				
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;				
OC	Hepacivirus.				
OX	NCBI_TaxID=11103;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=03249436; PubMed=7683463; "				
RA	Stuyver L., Arnhem W.V., Wyseur A., Deleys R., Maertens G.;				
RT	"Analysis of the putative E1 envelope and NS4a epitope region of HCV				
RT	type 3.";				
RL	Biochem. Biophys. Res. Commun. 192:635-641(1993).				
DR	EMBL; D14600; BAA03449.1; "				
DR	HSSP; P27958; IHEI.				
DR	MEROPS; S29.001; "				
DR	MEROPS; U39.001; "				
DR	InterPro; IPR000745; HCV_NS4a.				
DR	Pfam; PF01006; HCV_NS4a; 1.				
FT	NON_TER	1			
FT	NON_TER	209	209		
SQ	SEQUENCE	209 AA;	23408 MW;	76648D9BB1D3CD12 CRC64;	
Query Match					
Best Local Similarity 98.9%; Score 691; DB 12; Length 209;					
Matches 133; Conservative 0; Mismatches 1; Indels 0; Gaps 0;					
QY	1	QNEICLTHPTIKYIMACMSADLEVTSTVLLGGVLAALAAAYCLSGCVVIVGHIELGK 60			
Db	77	QNEICLTHPTIKYIMACMSADLEVTSTVLLGGVLAALAAAYCLSGCVVIVGHIELGK 136			
QY	61	PAIVDPKEVLYQQYDEMECSQAAYIEQAQVIAHQPKGVGLGLQRATQQAAVIEPIVT 120			

```
Db 137 PAIVDPKEVLYQYDEMEECQAAPYIEQAQVIAHQFKVGLGLQRATQQAIVPIVT 196
      |||
QY 121 TNWQKLEAFWKKH 133
      |||
Db 197 TNWQKLEAFWKKH 209
      |||

RESULT 2
O92933
ID O92933 PRELIMINARY; PRT: 3021 AA.
AC O92933;
DT 01-NOV-1998 (TRENBLrel. 08, Created)
DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-type 3a;
RA Shukla D.B., Chaturvedi S., Cao J.Y., Hoynes P.A.;
RT Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RL "Complete Nucleotide Sequence of the genome of Hepatitis C Virus type
CC 3a (CB).";
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL: AF046866; AAC03058.1; -.
DR HSSP: P27958; 1HEI.
DR MEROPS: S29.001; -.
DR MEROPS: U39.001; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002522; HCV_capsid.
DR InterPro: IPR002521; HCV_core.
DR InterPro: IPR002519; HCV_env.
DR InterPro: IPR002531; HCV_NS1.
DR InterPro: IPR002518; HCV_NS2.
DR InterPro: IPR004109; HCV_NS3.
DR InterPro: IPR000745; HCV_NS4a.
DR InterPro: IPR001490; HCV_NS4b.
DR InterPro: IPR002868; HCV_NS5a.
DR InterPro: IPR002166; HCV_RDRP.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR Pfam: PF01543; HCV_capsid; 1.
DR Pfam: PF01542; HCV_core; 1.
DR Pfam: PF01539; HCV_env; 1.
DR Pfam: PF01538; HCV_NS1; 1.
DR Pfam: PF01538; HCV_NS2; 1.
DR Pfam: PF02907; HCV_NS3; 1.
DR Pfam: PF01006; HCV_NS4a; 1.
DR Pfam: PF01001; HCV_NS4b; 1.
DR Pfam: PF01506; HCV_NS5a; 1.
DR Pfam: PF00998; Viral_RDRP; 1.
DR ProDom: PD186062; HCV_NS1; 1.
DR SMART: SM00487; DEXDC; 1.
DR PROSITE: PS05057; RDRP_POSITIVE; 1.
DR PROSITE: PS0521; RDRP_VIRAL; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3021 AA; 328903 MW; D7B6133B330303CD CRC64;

Query Match 96.68; Score 675; DB 12; Length 3021;
Best Local Similarity 95.5; Pred. No. 1.9e-62;
Matches 127; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 1 QNEICLPHPTKTYMACMSADLEVTSTWVLLGVLAALAAAYCLSVGCVVIVGHIELGGK 60
      |||
Db 1638 QNEICLPHPTKTYMACMSADLEVTSTWVLLGVLAALAAAYCLSVGCVVIVGHIELGGK 1697
```

```
FT CHAIN 2431 3021 NS5B PROTEIN.
SQ SEQUENCE 3021 AA; 329092 MW; BF2B499AA55A58CF CRC64;

Query Match 96.4%; Score 674; DB 12; Length 3021;
Best Local Similarity 95.5%; Pred. No. 2.4e-62;
Matches 127; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 QNEICLTHPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
   ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1638 QNETCLTHPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 1697
   ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 61 PAIVPDKEVLYQYDEMECSQAAPYIEQAQVIAHQFKGVGLGLLQRTAQQAQVIEPIVT 120
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1698 PALVPDKEVLYQYDEMECSQAAPYIEQAQVIAHQFKGVGLGLLQRTAQQAQVIEPIVA 1757
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 121 TNNQKLEAFWKKH 133
   ||| |||||:|||||
Db 1758 TNNQKLEAFWKKH 1770
   ||| |||||:|||||

RESULT 4
Q81258 PRELIMINARY; PRT; 3021 AA.
AC Q81258;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NZL1;
RA Okamoto H.;
RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=NZL1;
RA Sakamoto M.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA (BY SIMILARITY).
DR EMBL; D17763; BAA04609.1; -.
DR HSP; P27958; IHEI.
DR MEROPS; S29.001; -.
DR MEROPS; U39.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR002016; Peroxidase.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
```

```
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS05007; RDRP_POSITIVE; 1.
DR PROSITE; PS05021; RDRP_VIRAL; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
KW Polyprotein; RNA-directed RNA polymerase; Transferrase; Transmembrane.
FT CHAIN 1 191 C.
FT CHAIN 192 383 E1.
FT CHAIN 384 735 E2/NS1.
FT CHAIN 736 1012 NS2.
FT CHAIN 1013 1663 NS3.
FT CHAIN 1664 1717 NS4a.
FT CHAIN 1718 1978 NS4b.
FT CHAIN 1979 2430 NS5a.
FT CHAIN 2431 3021 NS5b.
SQ SEQUENCE 3021 AA; 329574 MW; 38712CCBC0C19562 CRC64;

Query Match 96.3%; Score 673; DB 12; Length 3021;
Best Local Similarity 95.5%; Pred. No. 3e-62;
Matches 127; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 QNEICLTHPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
   ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1638 QNETCLTHPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 1697
   ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 61 PAIVPDKEVLYQYDEMECSQAAPYIEQAQVIAHQFKGVGLGLLQRTAQQAQVIEPIVT 120
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1698 PALVPDKEVLYQYDEMECSQAAPYIEQAQVIAHQFKGVGLGLLQRTAQQAQVIEPIVT 1757
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 121 TNNQKLEAFWKKH 133
   ||| |||||:|||||
Db 1758 TNNQKLEAFWKKH 1770
   ||| |||||:|||||

RESULT 5
Q81595 PRELIMINARY; PRT; 133 AA.
AC Q81595;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Nonstructural protein 4 (Fragment).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93249436; PubMed=7683463;
RA Stuyver L., Arnhem W.V., Wyseur A., Deleys R., Maertens G.;
RT "Analysis of the putative E1 envelope and NS4a epitope region of HCV
RT type 3.";
RL Biochem. Biophys. Res. Commun. 192:635-641(1993).
DR EMBL; D14602; BAA03451.1; -.
DR HSP; P27958; IHEI.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1 1
FT NON_TER 133 133
SQ SEQUENCE 133 AA; 14781 MW; 4BFF2128FD301691 CRC64;

Query Match 96.1%; Score 672; DB 12; Length 133;
Best Local Similarity 93.2%; Pred. No. 8.8e-64;
Matches 124; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 QNEICLTHPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 QNEICLTHPTTKYIMACMSADLEVTSTWVLLGGVLAALAAAYCLSVGCVVIVGHIELGGK 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 61 PAIVPDKEVLYQYDEMECSQAAPYIEQAQVIAHQFKGVGLGLLQRTAQQAQVIEPIVT 120
```

```
Db 61 PALVPDKEVLYQYDEMEECQAAPYIEQAQVIAHQFKELGLLQRTAQQAQVIEPTV 120
121 TNWOKLEAFWHKH 133
121 SNWOKLETFWHKH 133

RESULT 6
Q68233
ID Q68233 PRELIMINARY; PRT; 138 AA.
AC Q68233;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Nonstructural protein (Fragment).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3a;
RX MEDLINE=95146953; PubMed=7844535;
RA Greene W.K., Cheong M.K., Ng V., Yap K.W.;
RT "Prevalence of hepatitis C virus sequence variants in South-East
  Asia.";
RL J. Gen. Virol. 76:211-215(1995).
DR EMBL; U14269; AAC53958.1; -.
DR HSP; P27958; IHEI.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1
FT NON_TER 138
SQ SEQUENCE 138 AA; 15379 MW; 45236C0E5427B19F CRC64;

Query Match 95.3%; Score 666; DB 12; Length 138;
Best Local Similarity 93.2%; Pred. No. 4e-63;
Matches 124; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QNEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSVGCVWIVGHIELGK 60
Db 6 QNEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSVGCVWIVGHIELGK 65
Qy 61 PAIVPDKEVLYQYDEMEECQAAPYIEQAQVIAHQFKGVGLGLLQRTAQQAQVIEPTV 120
Db 66 PALVPDKEVLYQYDEMEECQAAPYIEQAQVIAHQFKGVGLGLLQRTAQQAQVIEPTV 125
Qy 121 TNWOKLEAFWHKH 133
Db 126 TNWOKLETFWHKH 138

RESULT 7
Q81495
ID Q81495 PRELIMINARY; PRT; 3021 AA.
AC Q81495;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=111103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K3a;
RA Date T.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=K3a;
RX MEDLINE=95053917; PubMed=7964640;
RA Yamada N., Manihara K., Mizokami M., Ohba K., Takada A., Tsutsumi M.,
RA Date T.;
RT "Full-length sequence of the genome of hepatitis C virus type 3a:
  comparative study with different genotypes.";
RL J. Gen. Virol. 75:3279-3284(1994).
CC -J- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
  LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
  PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
  PROTEIN C AND MRNA (BY SIMILARITY).
CC EMBL; D28917; BAA06044.1; -.
DR HSP; P27958; IHEI.
DR MEROPS; S29.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.
DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR ProDom; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXdc; 1.
DR PROSITE; PS05057; RDRP_POSITIVE; 1.
DR PROSITE; PS05052; RDRP_VIRAL; 1.
KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;
  Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3021 AA; 328386 MW; A97418FF36C062A4 CRC64;

Query Match 94.6%; Score 661; DB 12; Length 3021;
Best Local Similarity 93.2%; Pred. No. 5.8e-61;
Matches 124; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 1 QNEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSVGCVWIVGHIELGK 60
Db 1638 QNEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSVGCVWIVGHIELGK 1697
Qy 61 PAIVPDKEVLYQYDEMEECQAAPYIEQAQVIAHQFKGVGLGLLQRTAQQAQVIEPTV 120
Db 1698 PALVPDKEVLYQYDEMEECQAAPYIEQAQVIAHQFKGVGLGLLQRTAQQAQVIEPTV 1757
Qy 121 TNWOKLEAFWHKH 133
Db 1758 SNWOKLETFWHKH 1770

RESULT 8
Q68241
ID Q68241 PRELIMINARY; PRT; 138 AA.
AC Q68241;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Nonstructural protein (Fragment).
GN NS4.
OS Hepatitis C virus.
```



```
DE Nonstructural protein (Fragment).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1a;
RX MEDLINE=95146953; PubMed=7844535;
RA Greene W.K., Cheong M.K., Ng V., Yap K.W.;
RT "Prevalence of hepatitis C virus sequence variants in South-East Asia.";
RL J. Gen. Virol. 76:211-215(1995).
DR EMBL; U14259; AAC53948.1; -
DR HSP; P27958; IHEI.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01539; HCV_NS1; 1.
FT NON_TER 1
FT NON_TER 138
SQ SEQUENCE 138 AA; 15392 MW; F8CB866A53AA907B CRC64;

Query Match 78.1%; Score 546; DB 12; Length 138;
Best Local Similarity 76.7%; Pred. No. 2.5e-50;
Matches 102; Conservative 12; Mismatches 19; Indels 0; Gaps 0;

QY 1 QNEICLTHPTTKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSYGCVVIGHIELGGK 60
DB 6 QNEITLTHPTTKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSYGCVVIGHIELGGK 65
QY 61 PAIVPOKEVLYQYDEMEECSQAAPYIEQAQVIAHQFGKVLGGLQRATQQQAVIEPIVT 120
DB 66 PAIIPREVLYREFDEMEECSQHLPIYIEQGMMLAEQFKQKALGGLQTASQAQVIEPIVQ 125
QY 121 TNMQKLEAFWKKH 133
DB 126 TNMQKLEAFWAKH 138

RESULT 14
O36579 ID O36579 PRELIMINARY; PRT; 3011 AA.
AC O36579;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Genome polyprotein.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H77;
RX MEDLINE=97373636; PubMed=9228008;
RA Kolykhalov A.A., Agapov E.V., Blight K.J., Mihalik K., Feinstone S.M.,
RA Rice C.M.;
RT "transmission of hepatitis C by intrahepatic inoculation with transcribed RNA.";
RL Science 277:570-574(1997).
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).
DR EMBL; AF009606; AAB66324.1; -
DR HSP; P27958; IHEI.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002522; HCV_capsid.
DR InterPro; IPR002521; HCV_core.
DR InterPro; IPR002519; HCV_env.
DR InterPro; IPR002531; HCV_NS1.
DR InterPro; IPR002518; HCV_NS2.
DR InterPro; IPR004109; HCV_NS3.

DR InterPro; IPR000745; HCV_NS4a.
DR InterPro; IPR001490; HCV_NS4b.
DR InterPro; IPR002868; HCV_NS5a.
DR InterPro; IPR002166; HCV_RdRP.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF01543; HCV_capsid; 1.
DR Pfam; PF01542; HCV_core; 1.
DR Pfam; PF01539; HCV_env; 1.
DR Pfam; PF01560; HCV_NS1; 1.
DR Pfam; PF01538; HCV_NS2; 1.
DR Pfam; PF02907; HCV_NS3; 1.
DR Pfam; PF01006; HCV_NS4a; 1.
DR Pfam; PF01001; HCV_NS4b; 1.
DR Pfam; PF01506; HCV_NS5a; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00998; Viral_RdRP; 1.
DR PRODOM; PD186062; HCV_NS1; 1.
DR SMART; SM00487; DEXDC; 1.
DR PROSITE; PS05057; RDRP_POSITIVE; 1.
DR PROSITE; PS05021; RDRP_VIRAL; 1.
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;
KW Hydrolase; Nonstructural protein; Polyprotein;
KW RNA-directed RNA polymerase; Transferase; Transmembrane.
SQ SEQUENCE 3011 AA; 327182 MW; E2E0EB809C63C1B9 CRC64;

Query Match 77.1%; Score 539; DB 12; Length 3011;
Best Local Similarity 75.9%; Pred. No. 5.7e-48;
Matches 101; Conservative 10; Mismatches 22; Indels 0; Gaps 0;

QY 1 QNEICLTHPTTKYIMACMSADLEVTSTWVLGGVLAALAAAYCLSYGCVVIGHIELGGK 60
DB 1632 QNEVTLTHPTTKYIMTCMSADLEVTSTWVLGGVLAALAAAYCLSTGCYVIGHIELGGK 1691
QY 61 PAIVPOKEVLYQYDEMEECSQAAPYIEQAQVIAHQFGKVLGGLQRATQQQAVIEPIVT 120
DB 1692 PAIIPREVLYQEFDEMEECSQHLPIYIEQGMMLAEQFKQKALGGLQTASQAQVIEPIVQ 1751
QY 121 TNMQKLEAFWKKH 133
DB 1752 TNMQKLEAFWAKH 1764

RESULT 15
Q68224 ID Q68224 PRELIMINARY; PRT; 138 AA.
AC Q68224;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Nonstructural protein (Fragment).
GN NS4.
OS Hepatitis C virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Hepacivirus.
OX NCBI_TaxID=11103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1a;
RX MEDLINE=95146953; PubMed=7844535;
RA Greene W.K., Cheong M.K., Ng V., Yap K.W.;
RT "Prevalence of hepatitis C virus sequence variants in South-East Asia.";
RL J. Gen. Virol. 76:211-215(1995).
DR EMBL; U14260; AAC53949.1; -
DR HSP; P27958; IHEI.
DR InterPro; IPR000745; HCV_NS4a.
DR Pfam; PF01006; HCV_NS4a; 1.
FT NON_TER 1
FT NON_TER 138
SQ SEQUENCE 138 AA; 15295 MW; 01335FB49A841A53 CRC64;
```

```

Query Match      77.0%; Score 538; DB 12; Length 138;
Best Local Similarity 75.2%; Pred. No. 1.7e-49;
Matches 100; Conservative 12; Mismatches 21; Indels 0; Gaps 0;

QY      1 QNEICLTHPTIKYIMACMSADLEVTSTWLLGGVLAALAAAYCLSVGCVIIVGHTELGGK 60
      |||: ||||:||||| ||||| |||||:||||| ||||| ||||| ||||| ||||| |||||
Db      6 QNEVTLTHPTIKYINTCMSADLEVTSTWLVGGVLAALAAAYCLSTGCCVIVIGRIVLSGK 65

QY     61 PAIVPDKVLYQYDEMEECQAAPYIEQAQVIAHQFKGVGLGLQRATQQQAVIEPIVT 120
      |||: ||||:||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     66 PAIIPDREVLVREFDEMEECQHLPLYIEQGMWLAQFQKQKALGLLQTSRQAEVITPAVQ 125

QY     121 TNWQKLEAFWHKH 133
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     126 TNWQKLEAFWAKH 138
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Search completed: August 29, 2003, 11:16:56
Job time : 56 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: August 29, 2003, 11:25:05 ; Search time 1821 Seconds
(without alignments)
1775.121 Million cell updates/sec

Title: US-09-638-693-36
Perfect score: 699
Sequence: 1 ONEICLTHPTIKYIMACKSA.....VIEPIVTTNWKLEAFWHKH 133
Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0
Searched: 22781392 seqs, 12152238056 residues
Total number of hits satisfying chosen parameters: 45562784
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Command line parameters:
-MODEL=frame+p2n.model -DRV=xlh
-Q/cgn2_1/USPTO.spool/US09638693/runat_25082003_173259_21658/app_query.fasta_1.327
-DB-EST -QMr-fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosom62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09638693 @CGN_1.1.2810 @runat_25082003_173259_21658 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -FGAPOP=10 -FGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST.*
1: em_estba.*
2: em_esthum.*
3: em_estlin.*
4: em_estm.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hc.*
9: gb_est1.*
10: gb_est2.*
11: gb_hc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gsal.*

29: gb_gss2.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	85.5	12.2	379	12	BM305347	BM305347 EESTee29
2	85	12.2	954	13	BU522891	BU522891 AGENCOURT
3	84.5	12.1	648	12	BJ311781	BJ311781 AGENCOURT
4	83	11.9	358	13	BY203246	BY203246 BY203246
5	83	11.9	363	13	BY192117	BY192117 BY192117
6	83	11.9	366	13	BY175124	BY175124 BY175124
7	82.5	11.8	580	12	BJ311912	BJ311912 BJ311912
8	82.5	11.8	569	12	BJ300261	BJ300261 BJ300261
9	82.5	11.8	590	12	BJ305350	BJ305350 BJ305350
10	82.5	11.8	595	12	BJ315881	BJ315881 BJ315881
11	82.5	11.8	615	12	BJ282133	BJ282133 BJ282133
12	82.5	11.8	620	12	BJ301051	BJ301051 BJ301051
13	82.5	11.8	621	12	BJ305233	BJ305233 BJ305233
14	82.5	11.8	640	10	BE404536	BE404536 WHE0443_E
15	82.5	11.8	640	12	BJ306865	BJ306865 BJ306865
16	82.5	11.8	646	10	BF047840	BF047840 GC84a11.Y
17	82.5	11.8	649	12	BJ320695	BJ320695 BJ320695
18	82.5	11.8	712	12	BJ298192	BJ298192 BJ298192
19	82.5	11.8	842	13	BU913689	BU913689 AGENCOURT
20	82.5	11.8	899	14	CA788512	CA788512 AGENCOURT
21	82	11.7	354	13	BY232072	BY232072 BY232072
22	82	11.7	423	10	BB848935	BB848935 BB848935
23	82	11.7	766	10	BE569942	BE569942 601332153
24	81.5	11.7	325	13	BY181416	BY181416 BY181416
25	81.5	11.7	556	12	BJ287254	BJ287254 BJ287254
26	81.5	11.7	700	12	BJ321356	BJ321356 BJ321356
27	81.5	11.7	830	28	AZ199860	AZ199860 SP_1040_B
28	81	11.6	347	13	BY185818	BY185818 BY185818
29	81	11.6	347	13	BY185899	BY185899 BY185899
30	80.5	11.5	263	10	BE157861	BE157861 WR2-HT037
31	80.5	11.5	912	28	AZ138865	AZ138865 SP_0172_A
32	80.5	11.3	1021	14	CD256199	CD256199 AGENCOURT
33	79	11.3	319	10	BB453355	BB453355 BB453355
34	79	11.3	369	13	BY230456	BY230456 BY230456
35	79	11.3	487	10	BE496123	BE496123 WHE1261_B
36	79	11.3	541	12	BI111572	BI111572 602895931
37	79	11.3	595	10	BG590945	BG590945 EST488787
38	79	11.3	735	12	BG964191	BG964191 602828967
39	79	11.3	749	10	BF982976	BF982976 602306383
40	79	11.3	902	10	BF585252	BF585252 602101913
41	79	11.3	1102	29	CNS0683B	AL411453 T7 end of
42	78.5	11.2	550	12	BJ312127	BJ312127 BJ312127
43	78.5	11.2	641	28	AZ247332	AZ247332 RPCI-23-3
44	78.5	11.2	880	10	BF676586	BF676586 602086313
45	78.5	11.2	993	14	CB563406	CB563406 AGENCOURT

ALIGNMENTS

RESULT 1
BM305347
LOCUS BM305347 379 bp mRNA linear EST 02-JAN-2002
DEFINITION BTEESTee29g11.y1 Eimeria tenella M5-6 cDNA Neg Selected Eimeria tenella cDNA 5', mRNA sequence.
ACCESSION BM305347
VERSION BM305347.1 GI:18037053
KEYWORDS EST.
SOURCE Eimeria tenella
ORGANISM Eimeria tenella
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae; Eimeria.
REFERENCE 1 (bases 1 to 379)

+

US-09-638-693-36 (1-133) x BY192117 (1-363)

Db 113 TACAGGACTGGCGAGGCTGGCTCAACTGCGGTCCCTGGAATC 72

dendritic cells Mus musculus cDNA clone F630008I05 5' mRNA

ORGANISM MUS MUSCULUS

H., Naqashima, T., Numata, K., Okido, T., Pavan, W. J., Perteau, G.,

Fax: 81-45-503-9216

10 (11), 1757-1771 (2000)

further details.

92 a	104 c	102 q	68 t
------	-------	-------	------

Best Local Similarity: 27.66%

```

Query Match: 11.87% Indels: 16
DB: 13 Gaps: 3

US-09-638-693-36 (1-133) x BJ175124 (1-366)
QY 10 IleThrIysTyrIleMetAlaCysMetSerAla----- 20
DB 352 TTGTCTCATCTTCAAGGGGTGATTTCAGCTTAAAGTCTTGTTCATCATCTGT 293
QY 21 -----AspLeuGluValThrSerThrValLeuLeuGlyGlyValLeu 36
DB 292 AGCACATCCAGACTTAACCTGTGCGGCTGGAGTACTGAGGGGTGTCAA 233
QY 37 AlaAlaLeuAlaTyrCysLeuSerValGlyCysValValIleValGlyHisIleGlu 56
DB 232 CAGATGATCTCTGGCCAACTGCTGGGCGATCTGCTCATTTGGTCTTGTCTTC 173
QY 57 LeuGly--GlyLysProAlaIleValProAspLysGluValLeuTyr-----GlnGln 73
DB 172 ATAAGCCAGGACGACCTGCTCTCGACCTGTGCGGAGATCTCTCCTCAGCAAGTGG 113
QY 74 TyrAspGluMetGluGluCysSerGlnAlaAlaProTyrIle 87
DB 112 TACAGGACTGCGGAGGCTGCTCAACTGCGGTCTCTCGAATC 71

RESULT 7
BJ311912
LOCUS BJ311912 560 bp mRNA linear EST 18-SEP-2002
DEFINITION aestivum cDNA clone whyd29015 3', mRNA sequence.
ACCESSION BJ311912
VERSION BJ311912.1 GI:23159531
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
TITLE Triticaceae; Triticum.
JOURNAL 1 (bases 1 to 560)
COMMENT Ogiwara,Y. and Murai,K.
Unpublished
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1. 560
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="whyd29015"
/tissue_type="spikelet at late flowering"
/dev_stage="Feekes' scale 6"
/clone_lib="Y. Ogiwara unpublished cDNA library, Wh_yd"

BASE COUNT 112 a 160 c 181 g 107 t
ORIGIN

Alignment Scores:
Pred. No.: 5.64 Length: 560
Score: 82.50 Matches: 38
Percent Similarity: 42.07% Conservative: 23
Best Local Similarity: 26.21% Mismatches: 55
Query Match: 11.80% Indels: 29
DB: 12 Gaps: 7

US-09-638-693-36 (1-133) x BJ311912 (1-560)
QY 7 ThrHisProIleThrIysTyrIleMetAlaCysMetSerAlaAspLeuGluValThrThr 26

```

```

126 ACCCATCAGCTGGAT-----CTAGGGGTCTAGAGACGATCTTTGATCTCTCGAC 176
QY 27 SerThrTrpValLeuLeuGly--GlyValLeuAlaAlaLeuAlaAlaTyrCysLeuSer 45
DB 177 GCCGTGCACATGTGCCGCGGACGGAGTGCTCGAAGCAGTGGCGCGCTGTTGCAGCC 236
QY 46 ValGlyCysValValIleVal-----Gly 53
DB 237 GCGGTGGGCGAGTGGTCTTTTCAGGGGTGCCCTGGGGGTGATGGAGACAGGATGGG 296
QY 54 HistLeuLeuGlyGly-----LysProAlaIleValProAspLysGlu 68
DB 297 CATGTGCGACTGGCGGTGCTACCTGTGTACACGCCCGCGCTGTACCGATGTGAA 356
QY 69 ValLeuTyrGlnGlnTyrAspGluMetGluCysSerGlnAlaAlaProTyrIleGlu 88
DB 357 CTCGCTCACCGCGCTCTGCACCTTCAC-----GCGCCCGCT---CCA 398
QY 89 GlnAlaGlnValIleAlaHisGlnPhelLysGlyLysValLeuGlyLeuLeuGlnArgAla 108
DB 399 CGAGCACACGCTGCTACGACAC-----GTGGTCCCGTGGCGGAGCAGGAGCCGGCT 452
QY 109 ThrGlnGlnAlaValIleGluProIleValThrAsnTrpGlnLysLeuGluAla 128
DB 453 GTAGAGGAGCTGCTGTGGCCAGCGCTCGATACCGAGGTAGAGTCTCTTGGCCGACAG 512
QY 129 PheTrpHisLysHis 133
DB 513 GTTCTGCACCGCTCAC 527

RESULT 8
BJ300261/c
LOCUS BJ300261 569 bp mRNA linear EST 18-SEP-2002
DEFINITION aestivum cDNA clone whyd14 5', mRNA sequence.
ACCESSION BJ300261
VERSION BJ300261.1 GI:23154813
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
TITLE Triticaceae; Triticum.
JOURNAL 1 (bases 1 to 569)
COMMENT Ogiwara,Y. and Murai,K.
Unpublished
Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1. 569
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="whyd14"
/tissue_type="spikelet at late flowering"
/dev_stage="Feekes' scale 6"
/clone_lib="Y. Ogiwara unpublished cDNA library, Wh_yd"

BASE COUNT 105 a 207 c 159 g 97 t
ORIGIN

Alignment Scores:
Pred. No.: 5.78 Length: 569
Score: 82.50 Matches: 38
Percent Similarity: 42.07% Conservative: 23
Best Local Similarity: 26.21% Mismatches: 55

```



```

Query Match: 11.80% Indels: 29
DB: 12 Gaps: 7

US-09-638-693-36 (1-133) x BJ300261 (1-569)

QY 7 ThrHisProlethrLysTyrlleMetAlaCysMetSerAlaAspLeuGluValThrThr 26
   ||||| : : : : : ||||| : : : : :
Db 552 ACCATCAGCTGGAT-----CTAGGGTTGTAGAACAGCATCTTGATCTCGTCGAC 502

QY 27 SerThrTrpValLeuLeuGly---GlyValLeuAlaAlaLeuAlaAlaTyrcysLeuSer 45
   : : : : : ||||| : : : : : ||||| : : : : :
Db 501 GCCTGGCACTTGCCTGGGAGCGGAGTCTCGAACAGTGGCCGCGCTGTGTCAGCC 442

QY 46 ValGlyCysValValIleVal-----Gly 53
   ||| : : : : :
Db 441 GCGCGTGGGCAGTGGTCTTCAGGGGTGGCGGTGGGTGATGAGACACGGATGGG 382

QY 54 HisIleGluLeuGlyGly-----LysProAlaIleValProAspLysGlu 68
   ||||| : : : : : ||||| : : : : :
Db 381 CATGTGCGACTGGCGGTGCTACCTGGTACCACGCCCGCGCTGCTGACCGATGTGAA 322

QY 69 ValLeuTyrcysValValIleVal-----Gly 53
   ||| : : : : :
Db 321 CTCGCTCACACCGCGTCTGTCACCTTCAC-----GCGCCCGCT---CCA 280

QY 89 GlnAlaGlnValIleAlaHisGlnPheLysGlyLysValLeuGlyLeuGlnArgAla 108
   : : : : : ||||| : : : : : ||||| : : : : :
Db 279 CGAGCACACGTCTGACGACAC-----GTGGGTGGCGGTGGCGGAGGAGCCAGGCT 226

QY 109 ThrGlnGlnAlaValIleGluProIleValThrThrAsnTrpGlnLysLeuGluAla 128
   : : : : : ||||| : : : : : ||||| : : : : :
Db 225 GTAGGAGGAGTCTGCTGTGGCCAGCGGCTCGATACCGAGGTAGAGGTCTTGGCGCAG 166

QY 129 PheTrpHisLysHis 133
   ||||| : : : : :
Db 165 GTTCTGCACCGCTCAC 151

RESULT 9
BJ305350/c
LOCUS
DEFINITION
  BJ305350 Y. Ogiwara unpublished cDNA library, wh_yd Triticum
  aestivum cDNA clone whyd29015 5', mRNA sequence.
ACCESSION
  BJ305350
VERSION
  BJ305350.1 GI:23156914
KEYWORDS
  EST.
SOURCE
  Triticum aestivum (bread wheat)
  Triticum aestivum
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
  ; Triticeae; Triticum.
  Ogiwara,Y. and Murai,K.
  Expressed genes in Triticum aestivum
  Unpublished
  Contact: Tadasu Shin-i
  Center For Genetic Resource Information
  National Institute of Genetics
  1111 Yata, Mishima, Shizuoka 411-8540, Japan
  Tel: 81-559-81-6856
  Fax: 81-559-81-6855
  Email: tshini@genes.nig.ac.jp.
  Location/Qualifiers
    1..590
      /organism="Triticum aestivum"
      /mol_type="mRNA"
      /cultivar="Chinese Spring"
      /db_xref="taxon:4565"
      /clone="whyd29015"
      /tissue_type="spikelet at late flowering"
      /dev_stage="Peekes' scale 6"
      /clone_lib="Y. Ogiwara unpublished cDNA library, wh_yd"

BASE COUNT
  106 a 199 c 169 g 116 t
ORIGIN
  106 a 199 c 169 g 116 t

Alignment Scores:
Pred. No.: 6.1 Length: 590
Score: 82.50 Matches: 38
Percent Similarity: 42.07% Conservative: 23
Best Local Similarity: 26.21% Mismatches: 55
Query Match: 11.80% Indels: 29
DB: 12 Gaps: 7

US-09-638-693-36 (1-133) x BJ305350 (1-590)

QY 7 ThrHisProlethrLysTyrlleMetAlaCysMetSerAlaAspLeuGluValThrThr 26
   ||||| : : : : : ||||| : : : : :
Db 482 ACCATCAGCTGGAT-----CTAGGGTTGTAGAACAGCATCTTGATCTCGTCGAC 432

QY 27 SerThrTrpValLeuLeuGly---GlyValLeuAlaAlaLeuAlaAlaTyrcysLeuSer 45
   : : : : : ||||| : : : : : ||||| : : : : :
Db 431 GCCTGGCACTTGCCTGGGAGCGGAGTCTCGAACAGTGGCCGCGCTGTGTCAGCC 372

QY 46 ValGlyCysValValIleVal-----Gly 53
   ||| : : : : :
Db 371 GCGCGTGGGCAGTGGTCTTCAGGGGTGGCGGTGGGTGATGAGACACGGATGGG 312

QY 54 HisIleGluLeuGlyGly-----LysProAlaIleValProAspLysGlu 68
   ||||| : : : : : ||||| : : : : :
Db 311 CATGTGCGACTGGCGGTGCTACCTGGTACCACGCCCGCGCTGCTGACCGATGTGAA 252

QY 69 ValLeuTyrcysValValIleVal-----Gly 53
   ||| : : : : :
Db 251 CTCGCTCACACCGCGTCTGTCACCTTCAC-----GCGCCCGCT---CCA 210

QY 89 GlnAlaGlnValIleAlaHisGlnPheLysGlyLysValLeuGlyLeuGlnArgAla 108
   : : : : : ||||| : : : : : ||||| : : : : :
Db 209 CGAGCACACGTCTGACGACAC-----GTGGGTGGCGGTGGCGGAGGAGCCAGGCT 156

QY 109 ThrGlnGlnAlaValIleGluProIleValThrThrAsnTrpGlnLysLeuGluAla 128
   : : : : : ||||| : : : : : ||||| : : : : :
Db 155 GTAGGAGGAGTCTGCTGTGGCCAGCGGCTCGATACCGAGGTAGAGGTCTTGGCGCAG 96

QY 129 PheTrpHisLysHis 133
   ||||| : : : : :
Db 95 GTTCTGCACCGCTCAC 81

RESULT 10
BJ315881/c
LOCUS
DEFINITION
  BJ315881 Y. Ogiwara unpublished cDNA library, wh_yf Triticum
  aestivum cDNA clone whyf20124 5', mRNA sequence.
ACCESSION
  BJ315881
VERSION
  BJ315881.1 GI:23161135
KEYWORDS
  EST.
SOURCE
  Triticum aestivum (bread wheat)
  Triticum aestivum
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
  ; Triticeae; Triticum.
  Ogiwara,Y. and Murai,K.
  Expressed genes in Triticum aestivum
  Unpublished
  Contact: Tadasu Shin-i
  Center For Genetic Resource Information
  National Institute of Genetics
  1111 Yata, Mishima, Shizuoka 411-8540, Japan
  Tel: 81-559-81-6856
  Fax: 81-559-81-6855
  Email: tshini@genes.nig.ac.jp.
  Location/Qualifiers
    1..595
      /organism="Triticum aestivum"
      /mol_type="mRNA"
      /cultivar="Chinese Spring"
      /db_xref="taxon:4565"
      /clone="whyf20124"
      /tissue_type="spikelet at late flowering"
      /dev_stage="Peekes' scale 6"
      /clone_lib="Y. Ogiwara unpublished cDNA library, wh_yf"

BASE COUNT
  106 a 199 c 169 g 116 t
ORIGIN
  106 a 199 c 169 g 116 t

```

```

/clone="whyf20i24"
/tissue_type="spikelet at early flowering"
/dev_stages="Feekes' scale 6"
/clone.lib="Y. Ogihara unpublished cDNA library, wh_yf"
BASE COUNT 110 a 210 c 169 g 105 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 6.18 Length: 595
Score: 82.50 Matches: 38
Percent Similarity: 42.07% Conservative: 23
Best Local Similarity: 26.21% Mismatches: 55
Query Match: 11.80% Indels: 29
DB: 12 Gaps: 7

US-09-638-693-36 (1-133) x BJ315881 (1-595)
Qy 7 ThrHisProfilThrLysTyriLeMetAlaCysMetSerAlaAspLeuGluValThrThr 26
Db 542 ACCCATCAGCTGGAT-----CTAGGGGTGTAGAGACGATCTTGTATCTCGTCGAC 492
Qy 27 SerThrTrpValLeuLeuGly---GlyValLeuAlaAlaLeuAlaAlaTyrcysLeuSer 45
Db 491 GCCGTGGACATGCGCGCGGACGAGTGCTCGAAGCAGTGCGCCCGCTGTTCGACGCC 432
Qy 46 ValGlyCysValValIleVal-----Gly 53
Db 431 GCGGTGGGCGAGTGGTCTTTTCAGGGGTGCGCGGTGTAGGAGACACGATGGG 372
Qy 54 HisLeuGluLeuGlyGly-----LysProAlaIleValProAspLysGlu 68
Db 371 CATGTGCACATGCGCGGTGCTCACCTGTGTACACGCCCGCGCTGTACCGATGTGAA 312
Qy 69 ValLeuTyriGlnTyriAspGluMetGluGluCysSerGlnAlaAlaProTyriLeuGlu 88
Db 311 CTCGTCCACAGGGGTCTGCACCTTCAC-----GCCGCCCGT---CCA 270
Qy 89 GlnAlaGlnValIleAlaHisGlnPheLysGlyLysValLeuGluGlnArgAla 108
Db 269 CGAGCACACGTCGTACGACAC-----GTGGTGCGCGTGGCGGAGCAGGAGCCAGGCT 216
Qy 109 ThrGlnGlnGlnAlaValIleGluProIleValThrThrAsnTrpGlnLysLeuGluAla 128
Db 215 GTAGAGGAGCTGTGTGTGGCCAGCGCTCGATACCGAGGTAGAGGTCTTGGCCGACAG 156
Qy 129 PheTrpHisLysHis 133
Db 155 GTTCTGCACCGTCAC 141

RESULT 11
BJ282133/c
LOCUS BJ282133 615 bp mRNA linear EST 18-SEP-2002
DEFINITION BJ282133 Y. Ogihara unpublished cDNA library, wh_r Triticum aestivum cDNA clone whr25e22 5', mRNA sequence.
ACCESSION BJ282133
VERSION BJ282133.1 GI:23145224
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
REFERENCE 1 (bases 1 to 615)
AUTHORS Ogihara, Y. and Murai, K.
TITLE Expressed genes in Triticum aestivum
JOURNAL Unpublished
COMMENT Contact: Tadasu Shin-i
Center for Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="whr25e22"
/tissue_type="root"
/dev_stages="Feekes' scale 1"
/clone.lib="Y. Ogihara unpublished cDNA library, wh_r"
BASE COUNT 109 a 210 c 176 g 120 t
ORIGIN

Alignment Scores:
Pred. No.: 6.5 Length: 615
Score: 82.50 Matches: 38
Percent Similarity: 42.07% Conservative: 23
Best Local Similarity: 26.21% Mismatches: 55
Query Match: 11.80% Indels: 29
DB: 12 Gaps: 7

US-09-638-693-36 (1-133) x BJ282133 (1-615)
Qy 7 ThrHisProfilThrLysTyriLeMetAlaCysMetSerAlaAspLeuGluValThrThr 26
Db 510 ACCCATCAGCTGGAT-----CTAGGGGTGTAGAGACGATCTTGTATCTCGTCGAC 460
Qy 27 SerThrTrpValLeuLeuGly---GlyValLeuAlaAlaLeuAlaAlaTyrcysLeuSer 45
Db 459 GCGGTGGACATGCGCGCGGACGAGTGCTCGAAGCAGTGCGCCCGCTGTTCGACGCC 400
Qy 46 ValGlyCysValValIleVal-----Gly 53
Db 399 GCGGTGGGCGAGTGGTCTTTTCAGGGGTGCGCGGTGTAGGAGACACGATGGG 340
Qy 54 HisLeuGluLeuGlyGly-----LysProAlaIleValProAspLysGlu 68
Db 339 CATGTGCACATGCGCGGTGCTCACCTGTGTACACGCCCGCGCTGTACCGATGTGAA 280
Qy 69 ValLeuTyriGlnTyriAspGluMetGluGluCysSerGlnAlaAlaProTyriLeuGlu 88
Db 279 CTCGTCCACAGGGGTCTGCACCTTCAC-----GCCGCCCGT---CCA 238
Qy 89 GlnAlaGlnValIleAlaHisGlnPheLysGlyLysValLeuGluGlnArgAla 108
Db 237 CGAGCACACGTCGTACGACAC-----GTGGTGCGCGTGGCGGAGCAGGAGCCAGGCT 184
Qy 109 ThrGlnGlnGlnAlaValIleGluProIleValThrThrAsnTrpGlnLysLeuGluAla 128
Db 183 GTAGAGGAGCTGTGTGTGGCCAGCGCTCGATACCGAGGTAGAGGTCTTGGCCGACAG 124
Qy 129 PheTrpHisLysHis 133
Db 123 GTTCTGCACCGTCAC 109

RESULT 12
BJ301051/c
LOCUS BJ301051 620 bp mRNA linear EST 18-SEP-2002
DEFINITION BJ301051 Y. Ogihara unpublished cDNA library, wh_yd Triticum aestivum cDNA clone whyd5m04 5', mRNA sequence.
ACCESSION BJ301051
VERSION BJ301051.1 GI:23155140
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
REFERENCE 1 (bases 1 to 620)
AUTHORS Ogihara, Y. and Murai, K.
TITLE Expressed genes in Triticum aestivum
JOURNAL Unpublished
COMMENT Contact: Tadasu Shin-i
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Triticum.

```


